Physics and Our Universe: How It All Works

Richard Wolfson, Ph.D.

THE
GREAT
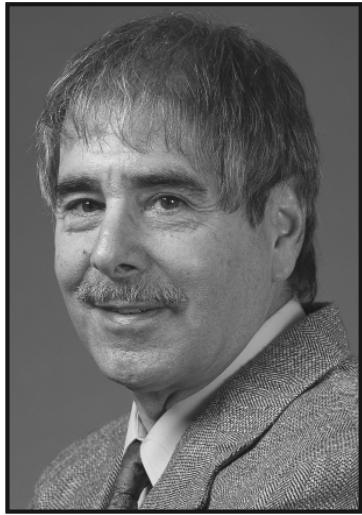COURSES®

Richard Wolfson, Ph.D.

Benjamin F. Wissler Professor of Physics
Middlebury College

Professor Richard Wolfson is the Benjamin F. Wissler Professor of Physics at Middlebury College, and he also teaches in Middlebury's Environmental Studies Program. He did undergraduate work at the Massachusetts Institute of Technology and Swarthmore College, graduating from Swarthmore with bachelor's degrees in Physics and Philosophy. He holds a master's degree in Environmental Studies from the University of Michigan and a doctorate in Physics from Dartmouth.

Professor Wolfson's books *Nuclear Choices: A Citizen's Guide to Nuclear Technology* (MIT Press, 1993) and *Simply Einstein: Relativity Demystified* (W. W. Norton, 2003) exemplify his interest in making science accessible to nonscientists. His textbooks include 3 editions of *Physics for Scientists and Engineers*, coauthored with Jay M. Pasachoff; 2 editions of *Essential University Physics* (Addison-Wesley, 2007, 2010); 2 editions of *Energy, Environment, and Climate* (W. W. Norton, 2008, 2012); and *Essential College Physics* (Addison-Wesley, 2010), coauthored with Andrew Rex. Professor Wolfson has also published in *Scientific American* and writes for *World Book Encyclopedia*.

Professor Wolfson's current research involves the eruptive behavior of the Sun's corona, as well as terrestrial climate change. His other published work encompasses such diverse fields as medical physics, plasma physics, solar energy engineering, electronic circuit design, nuclear issues, observational astronomy, and theoretical astrophysics.

In addition to *Physics and Our Universe: How It All Works*, Professor Wolfson has produced 3 other lecture series for The Great Courses, including *Einstein's Relativity and the Quantum Revolution: Modern Physics for Non-Scientists*, *Physics in Your Life*, and *Earth's Changing Climate*. He has

also lectured for the One Day University and *Scientific American*'s Bright Horizons cruises.

Professor Wolfson has spent sabbaticals at the National Center for Atmospheric Research, the University of St. Andrews, and Stanford University. In 2009, he was elected an American Physical Society Fellow. ∎

Table of Contents

Table of Contents

Table of Contents

Table of Contents

Table of Contents

Table of Contents

SUPPLEMENTAL MATERIAL

Physics and Our Universe: How It All Works

Scope:

Physics is the fundamental science. Its principles govern the workings of the universe at the most basic level and describe natural phenomena as well as the technologies that enable modern civilization. Physics is an experimental science that probes nature to discover its secrets, to refine our understanding, and to explore new and useful applications. It's also a quantitative science, written elegantly in the language of mathematics—a language that often permits us to predict and control the physical world with exquisite precision. Physics is a theoretical science, meaning that a few overarching "big ideas" provide solidly verified frameworks for explanation of broad ranges of seemingly disparate phenomena.

Our current understanding of physics traces to the work of Galileo and Newton in the 16th, 17th, and 18th centuries. Overthrowing 2000 years of misconceptions, these scientists laid the groundwork for the description of motion—a phenomenon at the heart of essentially everything that happens. The result is Newtonian mechanics: a simple, coherent theory expressed in 3 basic laws that even today describes most instances of motion we deal with in everyday life and, indeed, in much of the universe beyond Earth. Newtonian mechanics introduces some great ideas that continue throughout physics, even into realms where Newtonian ideas no longer apply. Concepts of force, energy, momentum, and conservation laws are central to all realms of physics—and all trace their origins to Newtonian mechanics. Galileo and Newton are also responsible for the first great unification in physics, as their ideas brought the terrestrial and celestial realms under a common set of physical laws. Newton's law of universal gravitation recognized that a universal attractive force, gravity, operates throughout the entire universe. Newton provided a mathematical description of that force, developed calculus to explore the ramifications of his idea, and showed definitively why the planets of our Solar System move as they do. Although Newtonian mechanics is more than 300 years old, it governs modern technologies ranging from skyscrapers to automobiles to spacecraft. This course begins, appropriately, with an exploration of Newtonian mechanics.

Motion manifests itself in more subtle ways than a car zooming down the highway or a planet orbiting the Sun. Wave motion transports energy but not matter; examples include ocean waves, seismic waves emanating from earthquakes, and sound. Liquids and gases, collectively called fluids, exhibit a wide range of motions, some strikingly beautiful and others—like the winds of a hurricane or the blast of a jet engine—awesomely powerful. Random motions of atoms and molecules are at the basis of thermodynamics, the science of heat and related phenomena. Thermodynamics governs many of the energy flows in the universe, from the outpouring of energy that lights the stars to Earth's complex climate system to the technologies we use to power modern society. Thermodynamics presents fundamental limitations on our ability to extract energy from fuels—limitations at the heart of today's energy concerns. Most phenomena of wave motion, fluid motion, and thermodynamics are ultimately explained in terms of Newtonian mechanics—a realization that gradually evolved in the centuries after Newton.

Electromagnetism is one of the fundamental forces in the universe and the dominant interaction on scales from atoms to our own bodies. Today, electrical and electronic technologies are indispensable; they range from the powerful motors that run our subways, high-speed railroads, and hybrid cars to the microchips that enable smart phones to have more computing power than the supercomputers of the late 20th century. Electromagnetism is also responsible for the forces that bind atoms into molecules and for molecular interactions that include, among many others, the replication of DNA allowing life to continue. Intimately related, electricity and magnetism together make possible electromagnetic waves. These waves provide nearly all the knowledge we have of the cosmos beyond our home planet, transport to Earth the solar energy that sustains life, and tie us increasingly to each other with a web of wireless communication—from traditional radio and television to cellular phone networks, GPS satellites, and wireless internet connectivity. As James Clerk Maxwell recognized in the mid-1800s, light is an electromagnetic wave—a realization that brought the science of optics under the umbrella of electromagnetism. This course devotes 12 lectures to electromagnetism.

Optics deals with the behavior of light. Phenomena of reflection, refraction, and interference are crucial to understanding and exploiting light. Eyeglasses, contact lenses, and laser vision correction all depend on optical principles—and so do the microscopes and telescopes that extend our vision to the interiors of living cells and to the most remote galaxies. DVD and Blu-ray discs store full-length movies in optically readable formats, and lasers exploit optics in applications from scanning barcodes to cutting metal. A total of 4 lectures explore optical principles and their applications.

Newtonian mechanics and electromagnetism comprise classical physics—a realm of physics whose theoretical background was in place before the year 1900 but that nevertheless remains relevant in much contemporary science and in many cutting-edge technologies. By 1900, physicists recognized seemingly subtle discrepancies between experimental results and classical physics. In the early decade of the 20th century, these discrepancies led to 2 revolutions in physics. Einstein's special and general theories of relativity radically altered our notions of space, time, and gravity. Quantum mechanics overthrew deep-seated classical ideas of determinism and causality. Together, relativity and quantum physics laid the groundwork for our modern understanding of the universe—the particles and fields that comprise it, the forces that bind components of it, and the interactions of those forces at the largest and smallest scales. This course ends with these revolutionary ideas and their applications today to cosmology, elementary particle physics, string theory, black holes, nanotechnology, and other topics at the cutting edge of modern physics. ■

The Fundamental Science
Lecture 1

P hysics is at the heart of our understanding of physical reality. Principles of physics apply universally—from the behavior of the infinitesimally tiny quarks that comprise the protons and neutrons of the atomic nucleus to the ordering of matter into galaxies, galaxy clusters, and superclusters at the largest scales imaginable. Physics also lays the groundwork for the other sciences, especially chemistry and biology. Nevertheless, emergent properties in complex systems mean that physics alone cannot provide a complete and comprehensible description of chemical and biological phenomena.

- Physics is the fundamental science; it's the most basic description we have of physical reality.

- Physics covers everything from the tiny subatomic particles called quarks and leptons to the stars, galaxies, clusters of galaxies, and large-scale structure of the entire universe itself.

- The interactions among the fundamental entities of physics give rise to the various scales of physics that are used to study the subject matter.

- The subatomic scale is the scale of elementary particles, such as the protons and neutrons that make up the nucleus of an atom. (Typical size: 1 femtometer, which is 10^{-15} meters, or 1/1,000,000,000,000,000.)

- The atomic and molecular scale is the scale of atoms and molecules. (Typical size: 10^{-9} meters, or 1 nanometer. The term "nano" means "1 billionth.")

- The nanotechnology scale is the scale of the smallest human-engineered structures, which have become so small that the scale is beginning to overlap with the atomic and molecular scale. (Typical size: between 1 and 100 nanometers.)

- The human scale is the scale of the everyday world. (Typical size: A typical human is somewhere between 1 and 2 meters tall, so it has a scale of about 1 meter.)

- The astronomical scale is the scale of planets, stars, and galaxies. (Typical size: a megameter, or 10^6 meters, to a zettameter. A zettameter is 10^{21} meters.)

- The cosmological scale is the scale of the largest things in the universe. (Typical size: 10^{26} meters.)

- If we understand how small-scale things in the universe—like atoms, molecules, quarks, and nucleons—work, we can determine how the large-scale things work as well.

- Surprisingly, what we now know about galaxies, clusters of galaxies, and the evolution of the universe also informs us about elementary particles and their interactions.

- One of the most fruitful things that has happened in physics in recent times is a symbiosis of **cosmology**, the study of the large-scale universe, and **particle physics**, the study of the very smallest particles in the universe.

Hubble data: NASA, ESA, and A. Zezas (Harvard-Smithsonian Center for Astrophysics); GALEX data: NASA, JPL–Caltech, GALEX Team, J. Huchra et al. (Harvard-Smithsonian Center for Astrophysics); Spitzer data: NASA/JPL/Caltech/ Harvard-Smithsonian Center for Astrophysics.

All matter—from human beings to galaxies—is composed of the same fundamental entities.

- The ultimate goal of physics is to determine a theory that would allow us to explain everything about the entire universe, sometimes called a **TOE**, or a **theory of everything**—but we aren't there yet.

- From the view of **reductionism**—a philosophical view that says we can reduce everything to basic physics—if we were able to understand the interactions of all the elementary particles, then we could understand everything there is to know about the universe.

- Most of the ordinary, everyday matter is composed of 3 elementary particles: an up quark, a down quark, and an electron.

- **Quarks** combine to make protons and neutrons, which are the building blocks of the nucleus of an atom, and electrons swarm around the nucleus.

- From chemistry, we know that atoms join to form molecules; however, physics helps us understand the physical principles that are used in the process.

- Cell biologists and microbiologists are learning rapidly what the molecular mechanisms are that allow cells to operate.

- Molecules join to form cells, which join to make organisms; somehow in quarks and electrons is the possibility of life.

- **Emergent properties** are properties of complicated systems, ultimately of the physical particles that make up the universe. For example, in physical systems, crystal structures are emergent properties; in biological systems, life itself is an emergent property.

- At some level of complexity, emergent properties become so interesting that, although we understand that they come from particles that are held together by the laws of physics, we can't understand or appreciate them through physics alone.

- Although physics is the fundamental science, once you recognize the existence of these emergent properties, you need the other sciences—the social sciences and history, for example—to understand them.

- In this course, we're going to use mathematics and demonstrations to understand the basic principles of physics—how they explain both natural and technological phenomena and, more importantly, how they lay the fundamental groundwork for our understanding of the entire universe.

- We're going to divide physics into a number of realms: classical mechanics (Newtonian mechanics), waves and fluids (the oscillation motion of fluids, gases, and liquids), thermodynamics (and statistical mechanics), electromagnetism (electricity and magnetism merged), optics (a branch of electromagnetism), and modern physics (the theory of relativity and quantum physics).

- The course is divided into 6 sections. The first section—Section 0 because it is an introduction to physics—consists of the first 2 lectures.

Important Terms

cosmology: The study of the overall structure and evolution of the universe.

emergent property: A higher-level property that arises from the micro-level interactions in a complex system.

particle physics: The study of the elementary constituents of nature.

quark: One of 6 fundamental particles with fractional charge that combine to make protons and neutrons, among other particles. The types include the up, down, charm, strange, top, and bottom quarks.

reductionism: The philosophical principle that complex systems can be understood once you know what they are made of and how the constituents interact.

theory of everything (TOE): The (as-yet hypothetical) theory that unites all known branches of physics, including classical mechanics, relativity, quantum theory, and so on.

Suggested Reading

Rex and Wolfson, *Essential College Physics* (ECP), chap 1.

Wolfson, *Essential University Physics* (EUP), chap 1.

Questions to Consider

1. Near the end of the 19th century, many scientists thought physics provided an essentially complete description of the principles underlying physical reality and that the future of physics would consist of merely exploring details and applications. (They were unimaginably wrong!) Today, some physicists are searching for a theory of everything, which would explain the entire physical universe in terms of a single interaction. Do you think such a theory would mark the end of physics as a grand quest for understanding and reduce it to the 19th-century expectation of merely exploring details and applications?

2. To what extent do you think the principles of physics underlie the sciences of chemistry and biology? To what extent do those sciences rely on their own principles, independent of physics principles?

Lecture 1: The Fundamental Science

Languages of Physics
Lecture 2

P hysics is mathematical, but its verbal language is just as important. Physicists give everyday words precise meanings, refining common usage or sometimes conveying altogether new concepts. Theories in physics provide the overarching conceptual framework for understanding vast realms of physical phenomena. The mathematics side of physics provides concise statements of physical principles that would be cumbersome to articulate using natural language. Mathematics expresses not only the numbers of physics, but also the relationships between physical quantities.

- Before delving into the actual physics, we're going to look at the languages of physics: the words and the mathematics we use to describe physical reality.

- Understanding the words and concepts used in physics—and understanding them precisely—is important.

- One of the reasons there are problems understanding the language of physics is because words are used slightly differently in science; scientific meanings of words might be related to everyday meanings, but they're more precise.

- For example, the everyday understanding of a theory is that it's a guess or hypothesis that might or might not be correct. In science, a **theory** is an overarching, coherent framework that explains and relates a whole body of scientific knowledge; it's been verified by many experiments and observations, and there are no internal contradictions or other experiments that have contradicted it.

- The public has a difficult time with uncertainty, so scientists have a difficult time communicating to the public sometimes because scientists—along with people in other areas human knowledge—are never 100% certain of anything.

- There are 3 reasons for the uncertainty in science, and the first is uncertainty of theories themselves: We think theories are right, but they might not be.

- Numerical uncertainty is next, and it takes several forms: the uncertainty in measurements (in which the measuring instruments are imperfect); the uncertainty in models; and the uncertainty in the projections of every physical, biological, and chemical event.

Courtesy NIST.

Iridium is one of the chemical elements used to construct the International Prototype Meter Bar.

- As we'll see when studying quantum physics, there is also a fundamental uncertainty described by something called the **Heisenberg uncertainty principle**, which makes uncertainty one of the fundamental properties of nature at the quantum (most fundamental) scale.

- Although mathematics is about numbers, it also expresses concisely profound and universal relations and interactions; we need numerical answers to answer quantitative questions.

- Numbers express the sizes of physical quantities; they have to be accurate and precise if they're going to be useful.

- Numbers also need to cover a very wide range of values, which is described by scientific notation. For example, 650,000,000 watts is 650 million watts or 650 megawatts, either of which can be written more easily as 6.5×10^8 W.

- Throughout the course, we will almost exclusively use the International System of Units (SI). For example, kilo stands for 1000, which means that a kilogram is 1000 grams and a kilometer is 1000 meters.

- On the other hand, we use the SI to measure small things: milli stands for thousandths, and micro, denoted by the Greek symbol mu (μ), stands for 10^{-6} or 1 millionth.

- In physics, most physical quantities have units, but the systems of units we use are human artifacts. It's important when you're doing a calculation or expressing a result to include the unit.

- Three of the fundamental physical quantities in mechanics are length, mass, and time. The meter is the SI unit of length, the kilogram is the unit of mass, and the second is the unit of time.

SI Prefixes		
Power	**Prefix**	**Symbol**
10^{24}	yotta	Y
10^{21}	zetta	Z
10^{18}	exa	E
10^{15}	peta	P
10^{12}	tera	T
10^{9}	giga	G
10^{6}	mega	M
10^{3}	kilo	k
10^{2}	hecto	h
10^{1}	deca	da
10^{0}	—	—
10^{-1}	deci	d
10^{-2}	centi	c
10^{-3}	milli	m
10^{-6}	micro	μ
10^{-9}	nano	n
10^{-12}	pico	p
10^{-15}	femto	f
10^{-18}	atto	a
10^{-21}	zepto	z
10^{-24}	yocto	y

- Each unit has a standard that has evolved over time, and the goal is to evolve toward a situation in which the standard of any one of these units is something any scientist anywhere could create or recreate in a laboratory.

- Mathematical equations express fundamental and sometimes crucial physical relations, and when an important equation comes up in this course, we're going to examine its anatomy by taking the equation apart and seeing how it works.

- Although you could go through this whole course without math, you will get more out of it if you follow the math.

- The expectations for your knowledge of mathematics in this course are high school algebra and trigonometry; the course only occasionally addresses calculus ideas.

- You should know how to solve simple equations from algebra: For example, if $2x = 6$, what is x? If you divide both sides by 2, you find that x is 3.

- Symbolically, if $ax = b$, what is x? By dividing both sides of the equation by a to get x alone, you find that x is b/a.

- **Trigonometry** is the study of the functions of angles—sine, cosine, and tangent—which are based on right triangles. The Greek symbol theta (θ) is the symbol commonly used for angles.

- In terms of general mathematics, you need to know how to read graphs, understanding what the axes mean in relation to one another.

Important Terms

Heisenberg uncertainty principle: The fundamental limit on the precision with which observers can simultaneously measure the position and velocity of a particle. If the position is measured precisely, the velocity will be poorly determined, and vice versa.

theory: A general principle that is widely accepted and is in accordance with observable facts and experimental data.

trigonometry: The branch of mathematics that studies the relationships among the parts of a triangle.

Lecture 2: Languages of Physics

Suggested Reading

Rex and Wolfson, *ECP*, chap 1.

Wolfson, *EUP*, chap 1.

Questions to Consider

1. Physics is the science that explains the physical universe in fundamental terms. But what about mathematics? Is it also a science grounded in physical reality, or is it more of a human construct?

2. The current unit of time, the second, is defined operationally in terms of the wavelength of light emitted by certain atoms. The current unit of mass is defined as the mass of the international prototype of the kilogram, kept at the International Bureau of Weights and Measures in Paris. Why is an operational definition, like that of the second, preferable to a standard object like the one that defines the kilogram?

3. Since 1983, it's been impossible to measure the speed of light. Why?

Describing Motion
Lecture 3

Without motion, the universe would be frozen in an instant of time. Motion is ubiquitous, from the swarming of electrons within an atom to the gravitational dance of distant galaxies. Fundamentally, motion is about changes in position. We quantify motion by first defining position and then introducing velocity as the rate of change of position. Velocity, too, can change, and its rate of change defines acceleration. Graphs of position, velocity, and acceleration versus time show how the 3 fundamental concepts of motion are related.

- In this first lecture in Section 1, we'll be describing motion without asking why motion occurs; we're simply going to talk about how we describe motion, both in words and in mathematics.

- Motion involves 2 important concepts: space (measured in meters, or m) and time (measured in seconds, or s).

- In this lecture and the next, we're going to simplify things by restricting motion to motion along a single line, which is called 1-dimensional motion.

- Example: From your house, you walk to an ice cream stand that's 1.2 kilometers (km) away at a steady speed and arrive in 20 minutes. It takes you 5 minutes to eat your ice cream, and then you head back at the same steady speed you walked before. Halfway home, you pause for 5 minutes to watch a construction project. You arrive home 50 minutes after you left.

- How far did you walk (distance)?

  ○ You walked 1.2 km to the ice cream stand and 1.2 km back, which is a total of 2.4 km.

- How fast did you walk (speed)?

  ○ On the way out, you went 1200 m (1.2 km), and it took 20 minutes, which is 1200 seconds. Speed is distance per time—1200 meters per 1200 seconds—which is 1 meter per second (m/s).

- What was your average speed for the entire trip (average speed)?

  ○ You walked a total of 2400 m, and you did it in 50 minutes, which is 3000 seconds. Your average speed is 2400 divided by 3000, or 0.8 m/s.

- How fast and in what direction were you going on your way to the ice cream stand (velocity)?

  ○ Your speed, as discovered previously, is 1 m/s, but we only know that you walked toward the ice cream stand.

- What overall change in your position occurred during the entire trip (displacement)?

  ○ Because you ended up back where you started, there was no change in your position, so your displacement is zero.

- Using these results, we can now develop some concepts about motion. We're going to use the variable x as a symbol for position, which describes where you are relative to some origin—arbitrarily described as position zero.

- In the case of motion in 1 dimension along a single line, positions can be thought of as lying along a number line; position is either positive or negative, depending in which direction an object travels from the origin.

- **Displacement**, a more precise concept that's related to **distance**, will be designated by the capital Greek letter delta next to the variable *x* (Δx).

- Throughout physics, the delta (Δ) symbol means a change in something; in this case, *x* stands for position, so Δx stands for the change in position.

- If you start at some position x_1 and you end up some position x_2, then your Δx is $x_2 - x_1$.

- Distance is about how much ground you've covered; it's always a positive quantity, and its direction is irrelevant.

- Because displacement is a change in position, the endpoints are all that matter; the details of the motion between those endpoints are completely irrelevant, and it can be positive or negative.

- In the example of the ice cream stand, the distance traversed was 2.4 km, but the displacement was zero.

- Speed and velocity are also related, but again, one is a physics concept and one is a more everyday concept.

- **Speed** is distance divided by time; it's always a positive quantity, and the direction of motion is irrelevant. Speed may vary depending on time interval.

- **Velocity** is displacement divided by time. Because displacement depends only on your starting and ending points, velocity (*v*) is Δx divided by Δt, the time interval involved in making that change in position.

- Again, the details of the motion are irrelevant, and displacement can be positive or negative. If the displacement is negative, then the velocity will also be negative, which tells you the direction of the velocity.

- The average speed over a time interval Δt is the total distance divided by the time. The average velocity is the displacement divided by the time, $\Delta x/\Delta t$.

- The instantaneous velocity is the result of taking that limit as Δt gets arbitrarily small (a concept of calculus). Instantaneous velocity is defined at every instant of time.

- As a function of time, instantaneous velocity may vary continuously with time, and it ends up being the slope of the position versus time graph.

- Instantaneous speed, then, is simply defined as the magnitude of the instantaneous velocity, in which the positive and negative values are irrelevant.

If you start at home and walk to an ice cream stand and back to your house, you have covered a distance, but there is no displacement because your initial and final positions are the same.

© Jupiterimages/liquidlibrary/© Getty Images/Thinkstock.

- Acceleration is analogous to velocity in the way that velocity is to position. **Acceleration** is the rate of change of velocity. The average acceleration is Δv, the change in velocity (or $v_2 - v_1$, the final velocity minus the initial velocity) divided by the time interval Δt.

- It is important to note that any change in velocity implies there's an acceleration.

- Instantaneous acceleration is similarly defined by taking the limit of an arbitrarily small Δt and becomes the slope of the velocity versus time graph.

- The unit of acceleration is meters per second per second (m/s/s) in the SI system, which is often written as meters per second, squared (m/s^2).

- The values of the motion concepts of position, velocity, and acceleration are not related; it's the rates of change that relate these 3 quantities, and they all may change continually with time.

Important Terms

acceleration: The rate of change of velocity, measured as distance divided by time.

displacement: The net change in position of an object from its initial to ending position.

distance: How far apart 2 objects are.

speed: The rate of change of position of an object, measured as distance over time.

velocity: Average velocity is total distance divided by the time it took to traverse that distance; units are length per time (for example, miles per hour).

Suggested Reading

Rex and Wolfson, *ECP*, chap 2.

Wolfson, *EUP*, chap 2.

Questions to Consider

1. The tip of your clock's minute hand is pointing straight up, at 12. An hour later it's again pointing at 12. Are its displacement and distance traveled during the hour the same? If not, explain the difference.

2. An advertisement boasts that a car can go from 0 to 60 mph in 7 seconds. What motion quantity is this ad describing?

3. Throw a ball straight up; at the top of its trajectory, it's instantaneously at rest. Does that mean it's not accelerating? Explain.

Lecture 3: Describing Motion

Falling Freely

Lecture 4

K nowing an object's initial position and velocity, and its subsequent acceleration, allows us to predict precisely its future position at any given time. This is the idea behind Newton's clockwork universe theory. In general, predicting the future position of an object presents a mathematical challenge that can be solved with calculus. But in the special case where acceleration doesn't change, a set of straightforward algebraic equations describes motion. Gravity near Earth's surface provides an important case of motion with constant acceleration.

- This lecture will take the motion concepts previously learned and use them to predict future motion.

- Inherent in Newtonian mechanics is the idea of the so-called clockwork universe in which the predictable universe unwinds like a giant clock.

- The idea of **determinism** is if we understand motion, then we understand how to predict future motion; however, that doesn't mean that all future motion is determined.

- This course deals only with the case of motion with **constant acceleration** (increases by the same amount over time) because it's easy to work with; motion with nonconstant acceleration is more difficult and generally involves calculus.

- This lecture uses equations to predict the velocity of an object undergoing constant acceleration.

- At some starting time (t), an object has some velocity (v_0) and moves with constant acceleration (a).

- Starting at the initial time, there's a time interval (t) over which the velocity changes from v_0 to some final value v.

- That change is the acceleration multiplied by the time (at) because acceleration is the rate of change of velocity.

- The velocity v is the initial velocity plus that change ($v = v_0 + at$). This is the prediction of the velocity at a future time.

- In this case of constant acceleration, the average velocity of an object between time 0 and time t is just half the starting velocity plus the final velocity ($v_{avg} = (1/2)(v_0 + v)$).

- In cases of **nonconstant acceleration**, the average velocity is not going to be halfway in between; it will be less or greater than a half, depending on the changes in acceleration and velocity.

- Now that we have its velocity, what is the position of the object as a function of time? Predicting its position at any future time is the basis of the clockwork universe idea.

The acceleration due to gravity near Earth's surface changes with an object's height.

- The equation to find an object's position (x) as a function of time (t) in the case of constant acceleration is $x_0 + v_0 t + (1/2)at^2$, where x_0 is the object's initial position, $v_0 t$ is the displacement (change in position) due to initial velocity, and $(1/2)at^2$ is displacement due to velocity resulting from acceleration.

- ○ Example: An airplane touches down at 80 m/s and then decelerates at a constant rate of 4 m/s². If the runway is only 900 m long, will the airplane stop before the end of the runway?

- ○ At the instant of touchdown, t is equal to 0. To make the math easier to work with, set the initial position (x_0) to 0.

- ○ If the positive direction is the direction toward the end of the runway, v_0 is positive 80 m/s. The acceleration (a) is −4 m/s²; the acceleration is opposite the velocity, indicating that the speed is decreasing (the plane is slowing down).

- ○ How long will it take the plane to stop? In other words, what will the time be when v is 0?

- Velocity with constant acceleration is the starting velocity plus the acceleration times time ($v = v_0 + at$).

- To solve the equation for time, insert 0 in for v to get $v_0 + at = 0$, which turns into $t = -v_0/a$. Insert 80 and −4 in for $-v_0$ and a, respectively, to discover that it will take 20 seconds for the plane to stop.

- But this doesn't answer our question: We want to know how far the plane will go in that 20 seconds. If x_0 is 0, v_0 is 80 m/s, a is −4 m/s², and t is 20 s. Plugging these numbers into the equation $x = x_0 + v_0 t + (1/2)at^2$ and solving for x, we find that the plane's position when it stops is 800 m—so it is safe by 100 m.

- The value of the acceleration near Earth's surface due to gravity is 9.8 m/s².

- In order to assume that this acceleration near Earth's surface is constant, we are making approximations by neglecting air resistance and the curvature of the planet, neither of which is perfectly true.

- Being in the state where the only force acting on you is gravity, you're experiencing gravity's acceleration in the state of **free fall**, which doesn't necessarily mean falling downward.

Important Terms

constant acceleration: Acceleration that increases by the same amount over time.

determinism: The belief that future events are completely determined by the present state of the universe—that is, by the exact positions and momenta of all of its particles.

free fall: A state in which gravity is the only force acting on an object.

nonconstant acceleration: Acceleration that does not increase by the same amount over time.

Suggested Reading

Rex and Wolfson, *ECP*, chap 2.

Wolfson, *EUP*, chap 2.

Questions to Consider

1. Two identical objects are dropped from rest, one from a building of height *h* and the other from a building of height 2*h*. Does the second object take twice as long to fall? Or does it take less than twice as long? Explain.

2. An example in this lecture calculated the minimum runway length for landing an airliner. If the airliner's landing speed were doubled, would the required runway length also double? If not, would it be less or more than twice the original length? What about the time for the plane to come to a stop? Explain your reasoning.

Lecture 4: Falling Freely

It's a 3-D World!
Lecture 5

We live in a 3-dimensional world; most significant cases of motion necessarily involve at least 2 and often all 3 dimensions. The richness of multiple dimensions requires a new mathematical tool for quantifying motion: the vector. A vector is a quantity that has both size—its magnitude—and direction. As in 1 dimension, the description of motion with constant acceleration in multiple dimensions is straightforward. Near Earth's surface, the result is so-called projectile motion, which provides a good approximation to the motions of objects.

- In the case of motion in 1 dimension, directionality is relevant only along a single line, back and forth—positive and negative.

- In 3 dimensions, it is possible to move in every direction, which expands the richness of what phenomena can happen but also makes the mathematics of describing motion more complicated.

- As we move into 3 dimensions, or even 2 dimensions, we have to introduce a new mathematical tool called the **vector**, a mathematical quantity that is deeper and richer than a number because it has both magnitude (how big it is) and direction.

- We can add, subtract, and multiply vectors by numbers and even by other vectors.

- The relationship in the 2-dimensional space between some origin (*O*) and some other point (*P*) can be described by a vector. The vector length is the distance from the origin *O* to the point *P*.

- The **paradigm vector** is a vector describing displacement.

- Displacement will be designated as Δr, while position measured with respect to some origin will be given by the vector **r**.

- If we add another point, Q, a different vector describes the displacement from P to Q—the vector $\Delta\mathbf{r}$, which is the change in position when you travel from P to Q.

Figure 5.1
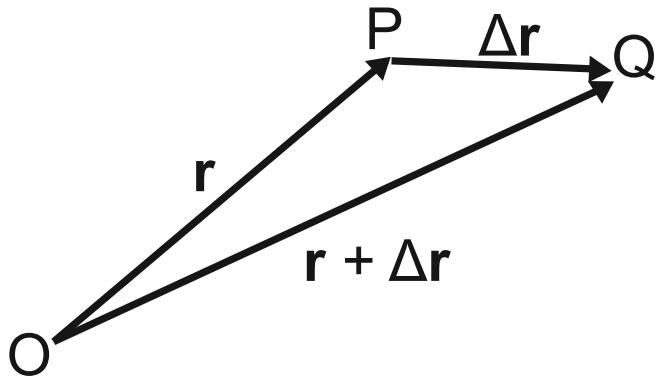
- The vector from O to Q describes a direct trip from O to Q, which is the sum of the vectors $\mathbf{r}$ and $\Delta\mathbf{r}$, or $\mathbf{r} + \Delta\mathbf{r}$.

- While a vector is a quantity with magnitude and direction, a number is a **scalar**, which is a quantity without direction.

- The vector $2\mathbf{v}$ is a vector in the same direction as $\mathbf{v}$, but it's twice as long; the vector $-\mathbf{v}$ is a vector of the same length as $\mathbf{v}$, but it's in the opposite direction.

- What's the vector $\mathbf{r}_2 - \mathbf{r}_1$? To get to the endpoint of $\mathbf{r}_2$ if you begin at the endpoint of $\mathbf{r}_1$, you have to add the difference of $\mathbf{r}_2 - \mathbf{r}_1$, or $\Delta\mathbf{r}$. In other words, the vector difference is the vector that has to be added to the second vector to get the first; it's also the vector achieved by adding the vector $\mathbf{r}_2$ and the vector $-\mathbf{r}_1$.

- Vectors are used in physics to represent quantities for describing motion. The **velocity vector** determines the rate of change of position.

- The **position vector** measures position with respect to some origin, or $\Delta\mathbf{r}$. The velocity vector is $\mathbf{v}$, or $\Delta\mathbf{r}/\Delta t$.

- The object's average velocity during a time interval Δt is the change in its position—its displacement, $\Delta\mathbf{r}$.

- Acceleration is the rate of change of velocity, and **a** is the acceleration vector.

- The object's average acceleration during a time interval (Δt) is found by taking the change in the object's velocity ($\Delta\mathbf{v}$)—the final velocity vector minus the initial velocity vector—divided by that time interval ($\mathbf{a} = (\mathbf{v} - \mathbf{v}_0)/\Delta t$).

- Any change in velocity represents an acceleration—a change in the magnitude of the velocity vector, a change in the direction of the velocity vector, or a change in both of them simultaneously.

- Acceleration and velocity do not have to have the same directions. Acceleration in the same direction as the velocity only serves to increase the speed, but acceleration opposite the velocity decreases the speed but doesn't change the direction.

- If the acceleration is at a right angle to the velocity, then the only thing that changes is the direction—but that's still an acceleration. In other words, you can be accelerating while your speed remains constant as long as your direction is changing.

- Projectile motion—the motion of an object moving under the influence of gravity—produces an arching path.

- Our coordinate system has a horizontal x-axis and a vertical y-axis, which is useful because the motion of an object in the horizontal direction is completely independent of the motion in the vertical direction.

- There are 2 equations we use to describe motion under the influence of gravity in 2 dimensions.

  ○ The first equation is the equation for position x in horizontal motion, but it has no acceleration term because gravity is only vertical. Its horizontal position (x) is the initial x position plus any initial x velocity ($x_0 + \mathbf{v}_{x_0} t$). The y is full accelerated motion with the constant acceleration of gravity; y is equal to the initial y position and the initial y velocity ($y_0 + \mathbf{v}_{y_0} t - 1/2gt^2$).

  ○ The second equation explains vertical position as a function of the horizontal position (trajectory). This equation describes the vertical position of an object starting at $x = 0$ and $y = 0$ with some initial speed $\mathbf{v}_0$ and at some angle θ to the horizontal.

Important Terms

paradigm vector: A vector that describes displacement.

position vector: A vector that describes position measured with respect to some origin.

scalar: A quantity without direction—just a number.

vector: A quantity that has both magnitude and direction.

velocity vector: A vector that determines the rate of change of position.

Suggested Reading

Rex and Wolfson, *ECP*, chap 3.1–3.4.

Wolfson, *EUP*, chap 3.1–3.5.

Questions to Consider

1. We generally consider the acceleration due to gravity to be constant near Earth's surface. What 2 factors make this assumption only approximately correct?

2. A line drive in baseball follows an almost straight, horizontal path from the batter to the outfield. Is a truly horizontal line drive possible? Explain.

3. Ideally, the trajectory of a projectile is a perfectly symmetric parabola in which the rising segment of the path is the same shape as the falling segment. Do you think the trajectory would remain symmetric in the presence of significant air resistance? If not, how might it change?

Going in Circles
Lecture 6

M otion on a circular path is an important case of 2-dimensional accelerated motion. Even when speed is constant, the motion is nevertheless accelerated because the direction of motion is changing. A simple expression gives the acceleration in circular motion in terms of the speed and the radius of the circle and shows that the direction of the acceleration is centripetal. Even when speed is changing, description of circular motion is straightforward because centripetal acceleration and acceleration along the direction of motion are considered independently.

- Circular motion is one of the most important types of motion throughout the universe in technological devices, nature, atomic-sized systems, planetary systems, and galactic systems.

- Because circular motion's direction is continually changing, it is accelerated motion, even if circular motion occurs at a constant speed—called **uniform circular motion**.

- There is circular motion when a car turns around a curve in a road. The wheel is turning as well as the entire car as it rounds the bend.

- As the car turns the corner, it is accelerating, and the wheel is accelerated because of its spinning motion. The entire car is accelerated because its motion is not in a straight line; even if the driver is holding the speed constant at a certain rate of miles an hour, it is still accelerated motion.

- Circular motion is ubiquitous; it occurs regardless of whether an object is traveling in a complete circle or just part of a circle.

- Accelerated motion is measured as the rate of change of velocity over the time interval involved ($\Delta \mathbf{v}/\Delta t$). The SI unit associated with accelerated motion is m/s^2.

- Even if the number of meters per second, the speed, is itself not changing, we can still characterize the acceleration, which is defined as the rate of change of velocity—a vector quantity—not as the rate of change of speed.

- Because it is attached to the object in accelerated motion, the velocity vector is tangent to the circular path of the object.

- The direction of its velocity vector is continually changing, and any change in velocity—whether in direction or magnitude or both—is, by definition, an acceleration.

- Figure 6.1 illustrates the initial position and the position after a little time has passed of a ball traveling in circular motion.

Figure 6.1
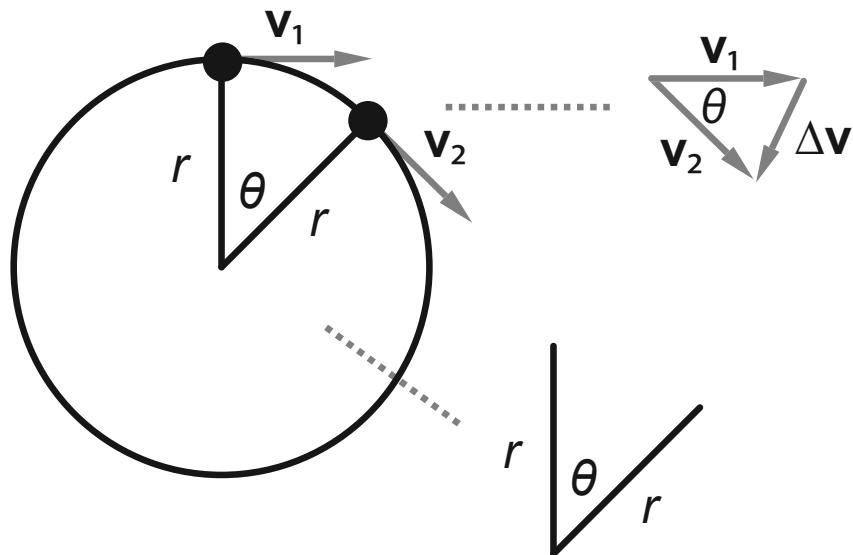
- The velocity vectors on each of the balls are exactly the same length. They are different velocities because they're pointing in different directions, but they represent motion with the same speed.

- The change in the ball's position is described by an angle designated by the Greek letter theta, θ.

- Because this is uniform circular motion in a perfectly circular path, the distance from the center of the ball's path is the same for both positions of the ball, designated by *r,* the radius of the circular path.

- The velocity vector is always tangent to the circle, which means that it is at right angles to the radius, from the center to the point where the ball is. The 2 velocity vectors are $\mathbf{v}_1$ and $\mathbf{v}_2$.

- Provided you don't change its length or orientation—that is, its direction—you can move a vector anywhere you want.

- The change between the 2 vectors ($\Delta\mathbf{v}$) is calculated by first putting $\mathbf{v}_2$ tail-to-tail with $\mathbf{v}_1$ and drawing a vector to connect them. The $\Delta\mathbf{v}$ needs to be added to the initial velocity to get the final velocity.

- As time goes on, the velocities are not the same; $\Delta\mathbf{v}$ is the measurable change between them. We need to divide that change by time ($\Delta\mathbf{v}/\Delta t$) to figure out exactly what the acceleration is in m/s^2.

- The acceleration is the change in velocity that occurred during the time interval, Δt, when the ball moved from its initial to its final position, which is approximately given by v^2, the square of the speed, divided by the radius of the circular path ($a = v^2/r$)—which is determined by using calculus.

- Because we know the direction of the acceleration is toward the center of the circle, we can call it **centripetal acceleration**, which means the acceleration points toward the center of the circle.

- With **nonuniform circular motion**, the speed of an object is changing even as it undergoes circular motion.

- Nonuniform circular motion is analyzed using what we understand of straight-line acceleration and acceleration in a circular path, which we now know has a magnitude of v^2/r.

Lecture 6: Going in Circles

- Nonuniform circular motion can be due to a changing path curvature or, possibly, to changing speed.

- Circular motion is never motion with constant acceleration. It may have constant speed, but it never has constant velocity.

- Never use a constant acceleration formula to describe circular motion; it has constant magnitude, but not constant direction.

Important Terms

centripetal acceleration: The acceleration of an object around any other object or position.

nonuniform circular motion: As an object undergoes circular motion, the speed of the object changes.

uniform circular motion: Circular motion that occurs at a constant speed.

Suggested Reading

Rex and Wolfson, *ECP*, chap 3.5.

Wolfson, *EUP*, chap 3.6.

Questions to Consider

1. You round a curve on the highway, and your speedometer reads a steady 50 mph. Are you accelerating? Explain.

2. The hour hand of a clock is half the length of the minute hand. Formulate an argument showing that the acceleration of the tip of the minute hand is 72 times greater than that of the hour hand's tip.

Causes of Motion
Lecture 7

O ver 2000 years ago, Aristotle incorrectly claimed that a force, a push or pull, was needed to sustain motion. Galileo and Newton got it right: They recognized that uniform motion—straight-line motion at constant speed—is a natural state that requires no explanation. What does require explanation is any change in motion. That's what forces do: They cause change in motion, which can be in speed, direction, or both. Newton summarized the role of force with his 3 laws of motion.

- The Greek philosopher scientist Aristotle explained that pushing or pulling something—called a **force**—causes motion.

- Aristotle's idea seems to make sense, but there are some areas of motion that are more difficult to explain with this idea. For example, if you shoot an arrow through the air, what causes the arrow to keep moving after the force from the bow stops acting? Aristotle said that air from the front of the arrow rushes in behind the arrow and pushes it to keep it moving.

- We can mathematically describe Aristotle's theory by saying that the velocity of an object, the rate of change of its position, is proportional to the force acting on it.

- The first person to really get a hint of how motion changed and why change was important in motion was Galileo, who did a number of **thought experiments**: experiments that are not actually conducted in the laboratory but have led to very powerful insights.

- From his thought experiments, Galileo concluded that the natural state for things to be in was a state of uniform motion, or moving at constant speed in a straight line.

- Through this reasoning, the **law of inertia** was formulated, which states that a body that's in uniform motion remains in motion, and a body that's at rest remains at rest, unless acted on by a nonzero net force.

- In mechanics, it is important to understand that while force causes motion to change, it's not the motion itself that requires force—it's the change in motion that requires force.

- Galileo set the stage for Isaac Newton, who became a codeveloper of calculus, which he needed for his equations of motion. He also worked on optics and developed his theory of gravity.

- Newton took Galileo's idea of inertia and used it to formulate 3 laws of motion.

  ○ Newton's first law simply restates the law of inertia.

  ○ Newton's second law quantifies the law of inertia by explaining how force and change in motion are related.

  ○ Newton's third law states that for every action, there is an equal and opposite reaction.

- Let's look at the second law in detail because the first law is a special case of the second law; the first law says if there's zero force, then there's zero change in motion.

- Newton quantified motion as the product of the object's mass and velocity, which is called **momentum**. He said that the rate of change of this quantity of motion is equal to the net force that's acting on the object.

- In the case of constant mass, Newton's second law reduces to the formula that's widely known as $\mathbf{F} = m\mathbf{a}$, force equals mass times acceleration.

- This formula is a restatement of Newton's original idea that the quantity of motion changes in proportion to the force in which the rate of change of an object's velocity is proportional to the net force and inversely proportional to its mass (both force and acceleration are vector quantities).

- Newton's second law says $\mathbf{F}_{net} = m\mathbf{a}$, in which $\mathbf{F}$ defines a force unit that is represented by the **newton** symbol (N) in the SI.

- If a 1-kg mass undergoes an acceleration of 1 m/s², then the net force acting on it is, by definition, 1 newton. (The English unit for mass is pounds, and 1 pound is 4.45 newtons.)

The Greek philosopher Aristotle discovered that forces cause motion.

© Photos.com/© Getty Images/Thinkstock.

- $\mathbf{F}_{net}$ is the **net force**, or the vector sum of all the different forces that may be acting on an object, and those are real physical forces.

- If there are 2 physical forces that point in the same direction, their vector sum, the net force, is simply a vector twice as long as either of them.

- If the 2 forces act in opposite directions and they're the same length—indicating that the forces have the same magnitude—they sum up to 0 net force.

- There also might be a situation in which a couple of forces are acting in different directions.

- There are common, everyday forces—called normal forces— that act at right angles to the surface they're in contact with. For example, the floor is pushing you up while gravity is pulling you down.

Lecture 7: Causes of Motion

34

- There are other kinds of forces that are invisible, such as gravity, friction, and electromagnetic force.

- Physicists have identified 3 fundamental forces in the universe that they believe will someday all be understood as aspects of one common force.

  ○ Electromagnetism and the so-called electroweak forces are often considered one category.

  ○ Then there's the strong force, which hypothetically could merge with the electroweak forces to form a grand unified force.

  ○ The force that we understand least is the force of gravity, and we don't really see yet how to combine it with the other forces, but if we did, we'd have a theory of everything.

Important Terms

force: The phenomenon that causes an object to accelerate.

law of inertia: Newton's first law of motion, which states that a body in motion (or at rest) remains in uniform motion (or at rest) unless a force acts on it.

momentum: The tendency of an object to remain rotating (angular momentum) or to remain in motion in a straight line (linear momentum). Momentum is one of the conserved quantities in nature; in a closed system, it remains unchanged.

net force: The sum of all forces acting on an object.

newton (N): In the International System of Units, the net force required to accelerate a mass of 1 kilogram at a rate of 1 meter per second, squared. It is named for English physicist and mathematician Sir Isaac Newton.

thought experiment: A highly idealized experiment used to illustrate physical principles.

Suggested Reading

Rex and Wolfson, *ECP*, chap 4.1–4.2.

Wolfson, *EUP*, chap 4.1–4.3.

Questions to Consider

1. Why is asking, "What causes motion?" the wrong question to ask? What's the right question?

2. You're pulling your suitcase across the airport floor, obviously exerting a force, and the suitcase is moving with constant speed and in a straight line. What can you conclude about the net force on the suitcase?

3. In what sense is Newton's first law really a special case of the second law?

Using Newton's Laws—1-D Motion
Lecture 8

N ewton's second law provides the link between force and acceleration. One important force is gravity, which, near Earth's surface, results in a force called weight that is proportional to an object's mass. Because weight and mass are proportional, all objects experience the same gravitational acceleration. If you weigh yourself in an elevator, your apparent weight will be greater than normal as the elevator accelerates upward and less than normal when it accelerates downward.

- Newton's laws are valid only for experimenters or observers who are in uniform motion—who are not accelerating. Those frames of reference of such observers are called **inertial reference frames** because in them the law of inertia holds.

- Earth is rotating slowly enough that, except in a few special circumstances, it is an approximation to an inertial reference frame.

- One of the simplest purposes of Newton's second law ($\mathbf{F} = m\mathbf{a}$) is to measure mass: If we know the force that's acting on an object and we know its acceleration or can measure its acceleration, we can then calculate its mass.

- What is **mass**? One definition of mass is it's the amount of matter in an object, but a better way to think of mass in the context of Newton's law is as the a measure of inertia—a measure of resistance to change in motion.

- **Weight**, by definition, is the force that gravity exerts on an object, and it is often confused with mass—a more fundamental property that the object owns by virtue of its inertia.

- Mass in the SI system is measured in kilograms (kg), but force is in newtons (N). In the English system, force is in pounds, and the English unit for mass is the slug, which is rarely used.

- One reason that mass and weight get confused is because of a remarkable fact: At a given point in space on Earth's surface, weight and mass are directly proportional, which means that an object that weighs twice as much as another also has twice as much mass.

- Weight is something that depends on where an object is because it depends on the strength of gravity (for example, the acceleration of gravity on Mars is only 3.7 N/kg as compared to 9.8 N/kg on Earth), but mass is something that is inherent to an object and doesn't change.

- Because mass and weight are directly proportional, all objects on Earth fall with the same acceleration, and in many examples that involve gravity, the mass will cancel out.

NASA.

In an orbiting spacecraft, astronauts are apparently weightless because they are accelerating at the same rate as everything around them. They are not actually weightless; the force of gravity in space is just weaker than it is on Earth.

- Astronauts in a space station exhibit **apparent weightlessness** because they and everything around them have the same acceleration. They are in a circular orbit around Earth that is changing, which means there must be a force on them: gravity, which is only slightly weaker at the space station than on Earth.

- Imagine that an elevator has some total mass: the elevator plus its passengers. It happens to be moving and accelerating upward. What is the cable tension?

  - The elevator has mass m, and it's accelerating upward (although our results will work even if it's accelerating downward.) The cable has to be stronger than it would need to be just to support the weight of the elevator because it's going to have to exert greater forces when the elevator is accelerating.

  - The physical forces acting on the elevator are the cable tension (designated as **T**), which is pulling upward, and gravity ($m\mathbf{g}$), which is pulling downward. Because the elevator is accelerating upward, the tension force has to be greater than the gravitational force.

- Let's rewrite Newton's second law ($\mathbf{F} = m\mathbf{a}$) as the vector equation $\mathbf{T} + m\mathbf{g} = m\mathbf{a}$.

  - We don't need to worry about which way the vectors are going; the vector **T** happens to point up, but all we need to know is that the sum of these 2 vectors equals the mass times the acceleration, so we can remove the vector notation.

  - In this case, T is upward, so let's say that means it's positive, but mg is downward, so it's negative: $T - mg = ma$. Using algebra to solve it, this equation becomes $T = ma + mg$.

- ○ If *a* is zero, then *ma* is also zero, so *T* is going to equal *mg*, which makes sense because if the elevator isn't accelerating, the tension on the cable is exactly equal to its weight—it's just supporting the weight of the elevator.

- ○ If *a* is greater than zero, that means its accelerating upward, so *T* is greater than *mg*, and the tension is greater than the weight. The tension needs to be greater than the weight to provide the extra force that accelerates the elevator.

- If *a* were –*g* (if the acceleration is downward and has magnitude *g*, the same as the gravitational acceleration), then *T* is zero—there's no cable tension. This is the case of free fall, which is what happens when there's no force acting but gravity.

- In this elevator problem, it's also the case that your weight as measured by a scale in an elevator will vary; the scale force may be different from your actual weight, causing your apparent weight to be different.

- When you're in an elevator that's accelerating upward, then you are also being accelerated upward, and there must be a net upward force on you. The 2 forces that are acting on you are the force of the scale pushing up on your feet and the force of gravity pulling down on you.

- In this case, the scale force has to be greater than the gravitational force, and because the scale reads that force, it will read a weight that's higher than your normal weight.

Important Terms

apparent weightlessness: The condition encountered in any freely falling reference frame, such as an orbiting spacecraft, in which all objects have the same acceleration and, thus, seem weightless relative to their local environment.

inertial reference frame: A perspective from which a person makes measurements in which Newton's first law holds.

mass: A measure of an object's material content or an object's tendency to resist an acceleration.

weight: The force of the gravitational pull on a mass.

Suggested Reading

Rex and Wolfson, *ECP*, chap 4.3.

Wolfson, *EUP*, chap 4.4–4.5.

Questions to Consider

1. Distinguish mass and weight.

2. Astronauts are often described as being "weightless" or "in zero gravity." Are these terms strictly correct? Why or why not?

3. You're standing on a scale in an elevator. Why, as the elevator starts moving upward, is the scale reading greater than your weight?

Action and Reaction
Lecture 9

Newton's third law is among the most widely misunderstood aspects of classical physics—in part because it's generally described in terms of the archaic phrase "action and reaction." Actually, the third law describes forces between interacting objects, asserting that the force one object exerts on another is equal and opposite the force that the second object exerts on the first. It does not stand as a separate feature of classical mechanics but rather goes hand-in-hand with the second law to provide a consistent description of motion.

- Newton's third law is one of the most commonly misunderstood aspects of Newtonian physics because of the archaic, misleading language that we tend to use in describing it: For every action there is an equal and opposite reaction. (We will use "force" instead of "action" and "reaction.")

- A modern version of the translation of the third law implies that there isn't an active agent and a passive agent; instead, there is mutuality (equal and opposite forces) between the 2 objects that are interacting.

- Newton's third law is required for a coherent, consistent description of motion. It is also the basis of a conservation law that is in the quantum realm and the realm of relativity.

- Example: You have a 1-kg block and a 2-kg block sitting on a frictionless surface, and they're touching. If you push on the 1-kg block with 6 N of force to the right, what force does the 2-kg block exert back on the 1-kg block?

- ○ The total mass of this system is 1 kg + 2 kg = 3 kg. Applying Newton's second law (**F** = *m***a**), **F** is 6 and *m* is 3, which means that **a** = 2. In other words, 3 kg × 2 m/s² = 6 N. The acceleration of the whole system (the pair of blocks) is **a** = 2 m/s².

- ○ Apply Newton's second law to just the 2-kg block. Solving **F** = *m***a** for **F**—in which *m* is 2 kg and **a** is 2 m/s²—the net force that is on the 2-kg block must be 4 N, and it's to the right.

- ○ Newton's third law, therefore, says that the 2-kg block exerts a leftward force of the same magnitude but in the opposite direction (4 N) back on the 1-kg block, which answers our original question.

The Horse and Cart Dilemma

In the horse and cart dilemma, the horse is pulling the cart and exerting a force on the cart. The cart is pulling back on the horse and exerting an equal but opposite force on the horse. How does the system ever get moving?

The paired forces associated with Newton's third law are never acting on the same object. The horse is exerting a force on the cart and the cart is exerting a force back on the horse, but those forces don't sum up to zero because they aren't acting on the same object.

Instead, what is the net force on the cart? If there's no friction, the net force is just the force from the horse. The horse is exerting a force on the ground with its feet, and the ground is exerting a force back on the horse; if that force is bigger than the force the cart is exerting on the horse, then the net force on the system is a force to the right, so the system moves.

- **Friction** is a force opposing the relative motion between any 2 surfaces in contact. Friction arises ultimately from electrical forces between atoms in the surfaces that are interacting and is approximately proportional to the normal force between the surfaces.

- The force of friction is equal to some constant number—mu (μ), the coefficient of friction, which is a property of those 2 surfaces—multiplied by the normal forces between the 2 surfaces ($F_f = \mu N$).

- Typically, the values of μ range from around zero to greater than 1. For example, a waxed ski on snow, where you don't want there to be a lot of friction, can be as low as 0.04, while the rubber on dry concrete, which is important for stopping a car, can be close to 1 at 0.8.

PHILOSOPHIÆ

NATURALIS

Brodie

PRINCIPIA

MATHEMATICA.

Autore *J S. NEWTON*, *Trin. Coll. Cantab. Soc.* Matheseos Professore *Lucasiano*, & Societatis Regalis Sodali.

IMPRIMATUR·

S. PEPYS, *Reg. Soc.* PRÆSES.

Julii 5. 1686.

R. ASTRON. SOC.

LONDINI,

Juſſu *Societatis Regiæ* ac Typis *Joſephi Streater*. Proſtat apud plures Bibliopolas. *Anno* MDCLXXXVII.

© Photos.com/© Getty Images/Thinkstock.

Sir Isaac Newton wrote *Mathematical Principles of Natural Philosophy* in 1687.

- There are 2 types of friction: kinetic friction and static friction. Kinetic friction, friction associated with objects that are actually moving relative to each other, is less than the force of static friction, which occurs when the objects are at rest relative to each other.

- Static is greater than kinetic friction because when the 2 objects are at rest with respect to each other, the bonds between the atoms and surfaces have had a chance to solidify.

- Friction is sometimes a nuisance: It slows things down and makes machinery inefficient. However, it's also essential in walking and driving and many other activities.

- For example, when you walk, the foot that is in contact with the ground is actually pushing backward against the ground while the ground is pushing the other direction (forward), and that is ultimately the force that propels you—again, Newton's third law.

Important Term

friction: A force that acts between 2 objects in contact, opposing any relative motion between them.

Suggested Reading

Rex and Wolfson, *ECP*, chap 4.2.

Wolfson, *EUP*, chap 4.6.

Questions to Consider

1. For every action there is an equal and opposite reaction. Explain this common statement in the context of Newton's third law and the concept of force.

2. When a tractor trailer starts up, the cab pulls on the trailer and the trailer pulls back on the cab with a force that is, according to Newton's third law, equal to the force of the cab on the trailer. So how can the rig ever get going?

3. You're pushing a trunk across the floor when a child hops onto the trunk. Why does it become harder to push?

Newton's Laws in 2 and 3 Dimensions
Lecture 10

F orce and acceleration are both vectors; they have direction as well as magnitude. Newton's second law shows that acceleration and force are proportional, which means that an object's acceleration—but not necessarily its velocity—is in the direction of the net force. Applying this idea to 2-dimensional motion resolves many seeming puzzles of the physical world. This lecture explores applications of Newton's laws in multidimensional motion and explores the common forces, including normal forces, tension forces, and friction.

- All the richness and glory of what Newton's second law has to say about motion gets revealed when we look at motion in 2 and 3 dimensions.

- How fast does a car need to be moving around a loop-the-loop track to make it all the way around the loop? The answer is the car's velocity vector and the associated magnitude of that velocity vector, its speed **v**.

  - Let's assume that the loop constitutes a circle of some radius *r*. In reality, loop-the-loop roller coasters are not designed as perfect circles. However, at the top of the arch is a curvature radius (*r*) that could represent the full circle.

- We know that the acceleration direction for circular motion is toward the center of the circle; in this case, the acceleration direction is downward at the top of the arch.

- Therefore, because the force and mass are vectors and are in the same direction, the net force has to be downward; the sum of all the physical forces acting on a car at the top of the track is downward.

- In applying Newton's second law, we must first identify the forces on the rollercoaster: gravity (*m***g**) and a normal force (**N**) from the track that is pushing up and is perpendicular to the track.

- We can now rewrite Newton's second law with those 2 physical forces identified: gravitational force plus normal force is equal to the mass times the acceleration (*m***g** + **N** = *m***a**).

Figure 10.1
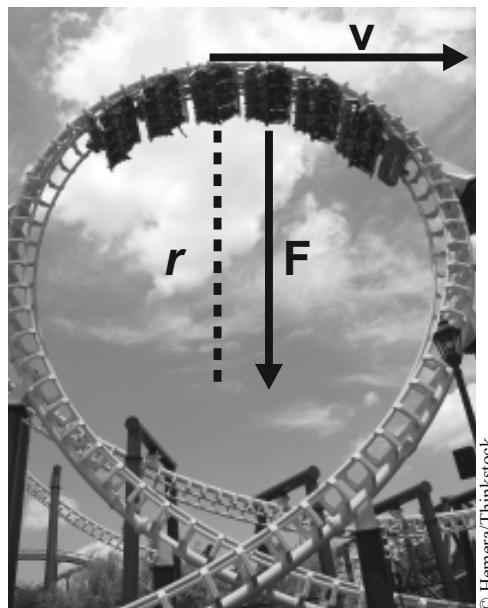
© Hemera/Thinkstock.

- For the car to make it around the loop, there must be a normal force between the car and the track. If that normal force were to reach zero, the car would at that instant lose contact and no longer be on the track.

- The top of the loop is the minimum condition for getting around the track, as the normal force just barely touches zero there. Then, as soon as the car gets passed the top, the normal force is going to grow again and be in good contact with the track. At that point, the only force acting is going to be gravity.

- At the top of the loop, Newton's law becomes *m***g** = *m***a**. Because we're talking about the direction of the net force being downward, we'll assume down to be positive.

- The acceleration in circular motion is the square of the speed divided by the radius, and as a vector, that acceleration is directed toward the center ($\mathbf{a} = \mathbf{v}^2/r$).

- Substitute $\mathbf{v}^2/r$ (the acceleration) into the previous equation $m\mathbf{g} + \mathbf{N} = m\mathbf{a}$ for $\mathbf{a}$, which results in the equation $mg - N = (mv^2)/r$. (We have removed the vector signs because we've assigned the convention that down is positive, and both the gravitational force and the acceleration are in the downward direction.)

- Solving for N, the equation becomes $N = m(g - v^2/r)$. The mass cancels because all objects experience the same gravitational acceleration regardless of their mass, so we get $v^2 = gr$. Dividing each side by r and taking the square root, $v = \sqrt{gr}$.

- Therefore, if you're going at speed $v = \sqrt{gr}$ at the top of the loop, you're going to just barely get around that loop.

- If you're going any faster, there will continue to be a nonzero normal force, and you'll remain on the track.

- If you're going any slower, you're going to lose contact with the track at the top of the trajectory and fall off in a parabolic arch like a projectile.

- What has to be the angle (θ) at which we bank a road in order for cars to be able to get around it without having to rely on friction to keep them on the road? In Figure 10.1, there is a racecar going around a banked track, which allows cars to have a normal force that has a component toward the center of the circle to keep them from sliding off the track.

    - The direction of the acceleration is toward the center of the turn (toward the left). The net force must be toward the center; it's horizontal.

Figure 10.2
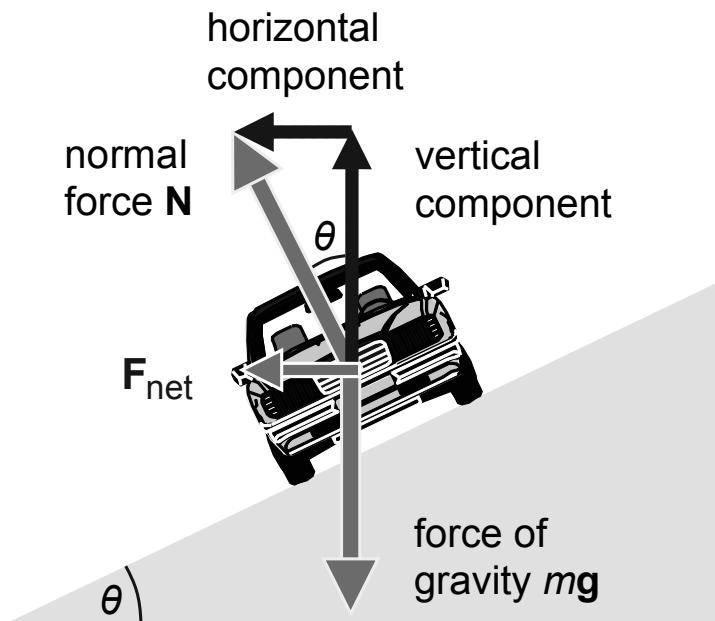
- What are the forces acting? They are the gravitational force $m\mathbf{g}$ and the normal force $\mathbf{N}$ in the road perpendicular to the road, which is not vertical—it's at the angle θ.

- If the car is not to be sliding along the road, the vertical component of the normal force has to balance gravity. The horizontal part, which is giving just the net force, has to be pointing to the left, and that's what is providing the acceleration of the car as it goes around the turn.

- Newton's second law for this case is $m\mathbf{g}$, the downward gravitational force, plus the normal force $\mathbf{N}$ equals $m\mathbf{a}$ ($m\mathbf{g} + \mathbf{N} = m\mathbf{a}$). (Let's take the convention that positive in the horizontal direction is toward the center of the circle, so we can remove the vector notation.)

- The acceleration in circular motion is the square of the velocity divided by the radius.

Figure 10.3

$$\sin\theta = \frac{\text{opposite}}{\text{hypotenuse}}$$

$$\cos\theta = \frac{\text{adjacent}}{\text{hypotenuse}}$$

$$\tan\theta = \frac{\text{opposite}}{\text{adjacent}}$$
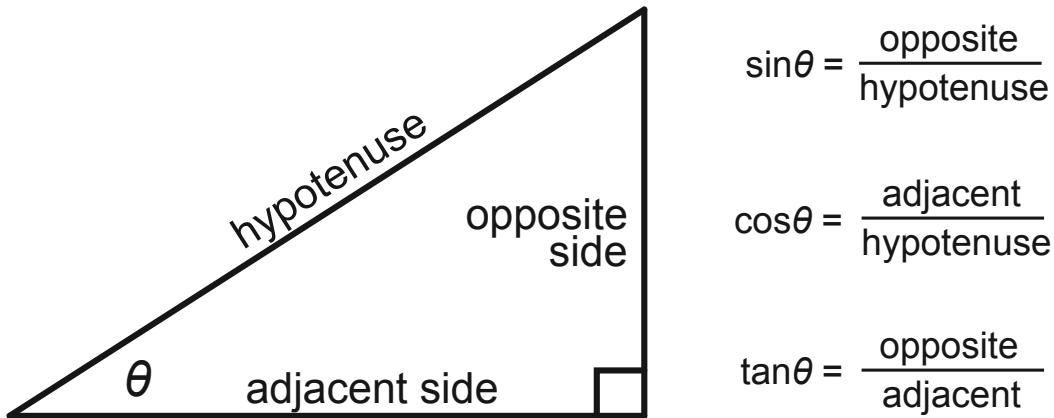
- We need trigonometry in this problem because we need to find out the 2 components of that normal force, horizontal and vertical. Because we are working with similar triangles, both components have the same angle θ.

- The horizontal component is the opposite, so the horizontal component over the normal force is the sine. The horizontal component of the normal force is $N\sin(\theta)$. There is no horizontal component of gravity, so $N\sin(\theta) = (mv^2)/r$.

- For the vertical component, let's use positive upward $-mg$ as the gravitational force plus $N\cos(\theta)$, the adjacent side, and set those equal to zero $(-mg + N\cos(\theta) = 0)$. Solving for N, $N = (mg)/\cos(\theta)$.

- Because we've solved 1 of these 2 equations for N, we can plug in $(mg)/\cos(\theta)$ (from the vertical component equation) for N in the horizontal component equation to form the equation $[(mg)\sin(\theta)]/\cos(\theta) = (mv^2)/r$. Solving this, we get $\tan(\theta) = v^2/(gr)$.

- As usual in problems involving gravity, the mass cancels, which means that the same banking angle holds for both a big truck and a small car.

- If you know the velocity you want cars to travel at for this turn and you know the radius of the turn, the answer for the banking angle you should make is described by $\tan(\theta) = v^2/(gr)$.

- If you build that banking angle, cars moving at your designated speed will not need frictional forces to hold them in their turn— they will be held there by the normal force alone.

Suggested Reading

Rex and Wolfson, *ECP*, chap 4.4–4.5.

Wolfson, *EUP*, chap 5.

Questions to Consider

1. What holds a satellite up? After all, gravity is pulling it toward Earth.

2. Why are highways banked?

3. A stunt pilot is flying in a circular path, oriented vertically. If the plane's speed is constant, at what point in the path does the seat exert the greatest force on the pilot? Explain.

Work and Energy
Lecture 11

I n physics, work results when a force acts on an object that moves, provided the force isn't at right angles to the object's motion. Energy is a central concept in physics; kinetic energy and potential energy are among the types of energy used in physics. One of the most important principles in physics is conservation of energy, which in this context states that as long as the only forces acting are conservative, then an object's total energy—the sum of its kinetic and potential energy—is unchanged.

- Work and energy have narrower meanings in physics than they do in everyday usage. **Work** is the product of the component of the force in the direction an object moves multiplied by the object's displacement.

- Work involves force and displacement, or force and motion. If an object doesn't move, you can be applying a force to it, but no work gets done.

- If the force is in the direction of the motion, then the work is just the product of the force with the distance the object moves; if the force is at some angle, you have to figure out the component of the force in the direction of the motion.

- **Energy** is the ability to do work. The unit for work and for energy is the **joule** (J)—named after James Joule, who discovered that heat is a kind of energy.

- In basic SI terms, a joule is a newton-meter (N·m), which is the amount of work done when applying a force of 1 newton to move an object a distance of 1 meter.

- Some examples of doing work are simple because the force is constant. Often, force varies with position in situations involving gravity and springs.

- Figure 11.1 shows a force that varies with position. Position is plotted on the horizontal axis, and the force, written $F(x)$ because it depends on position, is plotted on the vertical axis.

Figure 11.1
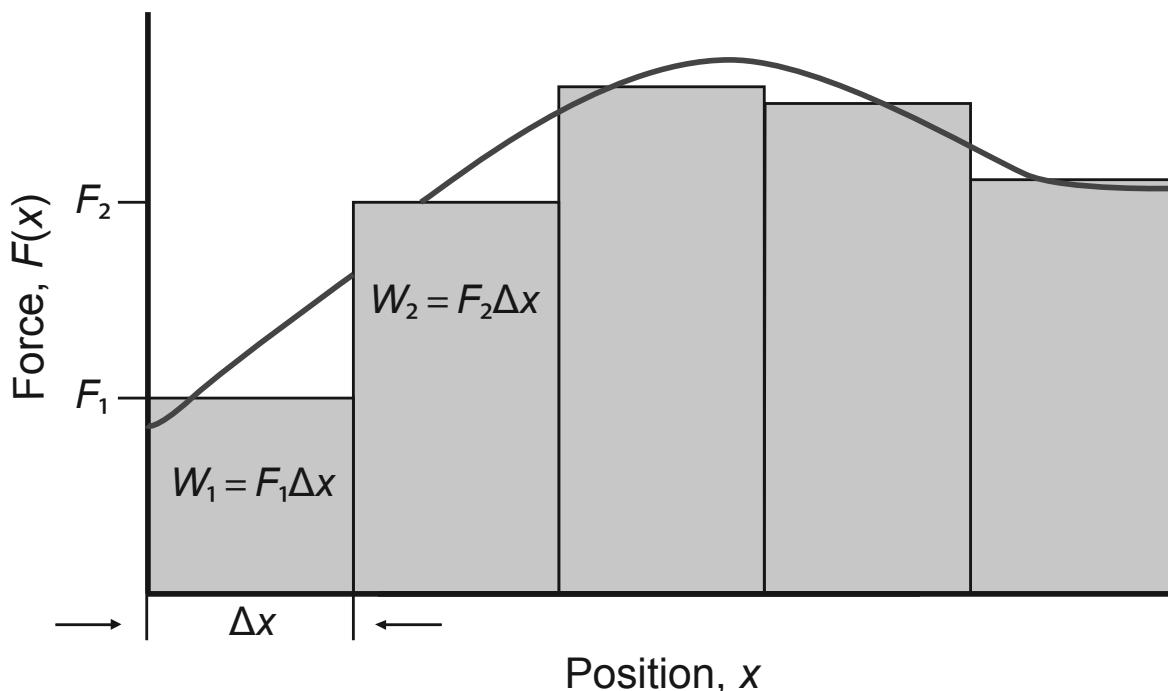
- For the interval of position Δx, we applied a force F_1, and therefore the work was approximately F_1 times that interval (displacement) Δx.

- This process can be repeated for F_2 and so on. The area of each rectangle, the force times the displacement, is the work from one position to the next, which can be added up to approximate the total amount of work done.

- If you use calculus, you will get a better answer, but either way, the work ends up being the area under the force-versus-position curve ($W = F_x \Delta x$).

- One particular example of a force that varies position is the force associated with springs, which stretch and exert forces. The more you stretch a spring, the greater the force.

- To quantify a spring's force, we introduce a quantity called k, the spring constant, which tells you how much force there will be for a given stretch.

- For stiff springs, a large force needs to be applied to stretch it even a small distance; a stiff spring has a much larger spring constant k than a limp spring.

- Figure 11.2 shows the force versus distance curve for an ideal spring, one that obeys what's called **Hooke's law**, in which the force of the spring is directly proportional to the stretch x.

- The area of the triangle, the area under the curve, is $(1/2)kx^2$—the work it takes to stretch a spring by a distance x from its equilibrium position.

- The work that is done when stretching a spring is somehow stored in that spring, and that stored work is called **potential energy** because it's not doing anything right now but has the potential to make a particular motion.

- There are 2 different kinds of forces: conservative forces and nonconservative forces.

  ○ Conservative forces give back the work that's done by pushing against them with forces. Gravity, electric forces, and elastic forces are examples of conservative forces. The work done against conservative forces is stored as potential energy.

Figure 11.2
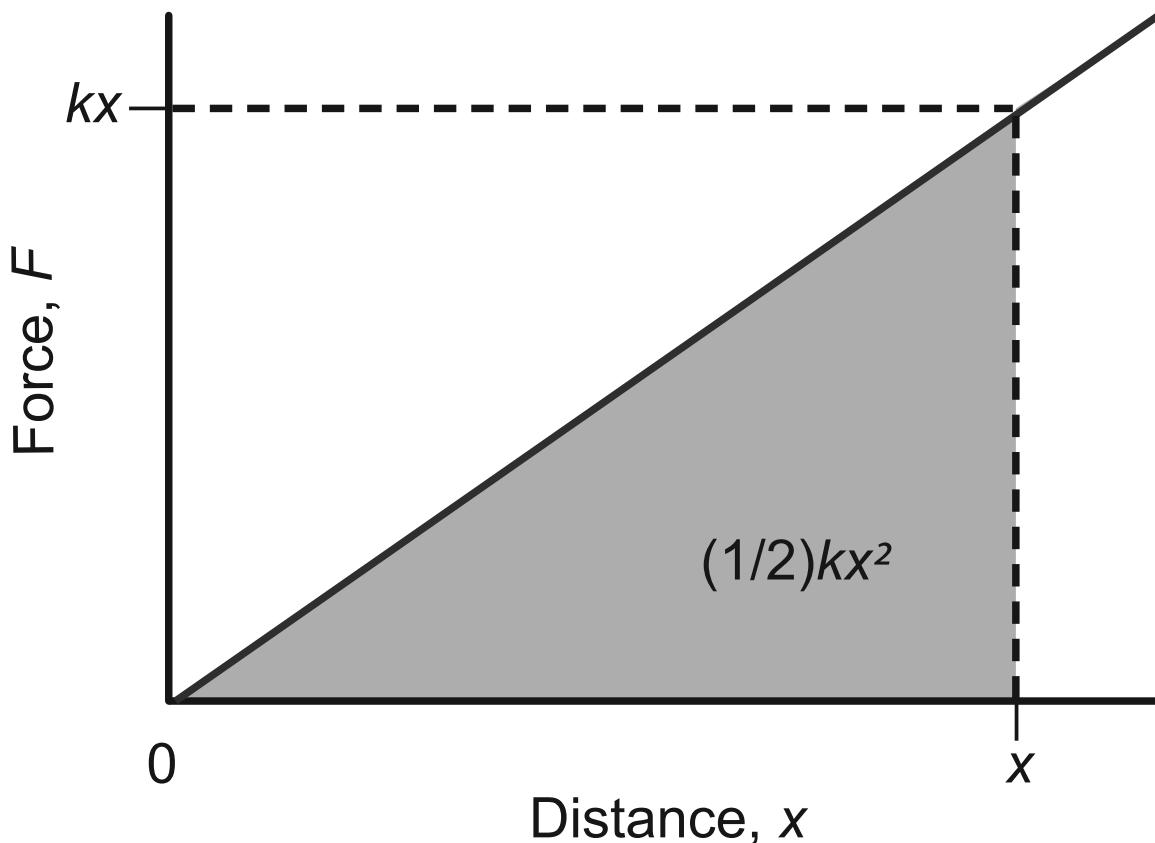
Work = Triangular area = $(1/2)kx^2$

- ○ Nonconservative forces don't give back the work that's done against them. Friction, air resistance, and the viscosity of fluid are examples of nonconservative forces, which produce heat rather than storing the work as potential energy.

- In the case of **gravitational potential energy**, if you lift a mass m a distance h, the mass has to apply a force equal to its weight mg, so the gravitational work done is mgh, which is the energy that then gets stored in the lifted object. We'll use the symbol U for potential energy, so $\Delta U_{gravitational} = mgh$.

- The **elastic potential energy**, the potential energy stored when stretching a spring, is $\Delta U_{elastic} = (1/2)kx^2$.

- We use ΔU for both equations because we arbitrarily define a zero of potential energy that is the equilibrium position, and we want to measure how much more potential energy was stored.

- In general, if you take the negative of the work done by conservative forces, you get the change in potential energy ($\Delta U = -W_{conservative}$). The negative represents the negative work done by the force against the work; for example, gravity pulls downward on an object that you are trying to move upward.

- Suppose you have an object subject to a constant net force, starting with speed v_1, moving a distance Δx, and ending at a speed v_2. The net amount of work that is done by the net force acting on the object is W ($W_{net} = F_{net}\Delta x$).

- Using Newton's second law, we find that $F_{net} = ma = m(\Delta v/\Delta t)$. Substituting this F_{net} into the net work equation, we get $W_{net} = [m(\Delta v/\Delta t)]\Delta x = (m\Delta v)(\Delta x/\Delta t)$.

- Substituting and using algebra, $W_{net} = m(v_2 - v_1)[(v_2 + v_1)/2]$. We can multiply this equation out to find that the net work involved is $(1/2)mv_2^2 - (1/2)mv_1^2$.

- This new quantity, $(1/2)mv^2$, is called **kinetic energy** (K) and changes only if net work gets done on an object. It is the energy associated with an object's motion and is written as $K = (1/2)mv^2$.

- To find this quantity, we applied Newton's second law to find the work done in moving an object with a net force as its velocity changed from v_1 to v_2.

- Although we assumed the force was constant to apply Newton's second law without using calculus, this is a general result called the **work-energy theorem**, which says that the change in kinetic energy is equal to the net work done on an object ($\Delta K = W_{net}$).

- The **law of conservation of energy**—the idea that energy can change form but can't disappear and so is conserved—is a principle that goes beyond Newtonian physics.

- If only conservative forces act on an object, then the net work is the work done by conservative forces. In that case, ΔK will be the work done by conservative forces, and that will be equal to minus the potential energy change ($\Delta K = -\Delta U$).

- Algebraically, we pull the ΔU to the other side where it becomes positive, and we have $\Delta K + \Delta U = 0$, which is the statement of conservation of mechanical energy. Therefore, mechanical energy is the sum of kinetic energy and potential energy.

- In a system in which only conservative forces act, the sum of the kinetic and potential energies doesn't change, but we can convert kinetic energy to potential and potential energy to kinetic as long as we keep the total energy constant.

Important Terms

elastic potential energy: The energy that is stored when stretching an object (a spring, for example), which can be measured with the equation $\Delta U_{elastic} = (1/2)kx^2$.

energy: The ability to do work.

gravitational potential energy: The energy content of an object due to its position in relation to other objects, which can be measured with the equation $\Delta U_{gravitational} = mgh$.

Hooke's law: In an ideal spring, the force of the spring is directly proportional to the stretch.

joule (J): In the International System of Units (SI), the amount of work that can be done by applying a force of 1 newton through a distance of 1 meter. It is named for 19th-century English physicist James Prescott Joule.

kinetic energy: The energy of motion. For a particle of mass m moving with velocity v, the kinetic energy is $K = (1/2)mv^2$.

law of conservation of energy: A fundamental law of physics that states that in a closed system, energy cannot be created or destroyed; it can only change form.

potential energy: The energy content that an object has by virtue of its chemical configuration or its position in space.

work: The exertion of a force over a distance.

work-energy theorem: The change in kinetic energy is equal to the net work done on an object ($\Delta K = W_{net}$).

Suggested Reading

Rex and Wolfson, *ECP*, chap 5.1–5.4.

Wolfson, *EUP*, chap 6.1–7.1.

Questions to Consider

1. You're impressed by a waitress carrying a heavy tray of food through a restaurant. If the restaurant floor is horizontal, is the waitress doing work in the physics sense of that word?

Lecture 11: Work and Energy

2. Distinguish conservative from nonconservative forces, and give an example of each.

3. You start with an unstretched spring and stretch it by 1 inch; in the process, you do a certain amount of work. If you stretch the spring another inch, do you do the same amount of work again, or do you do more or less work than with the first inch?

Using Energy Conservation
Lecture 12

The law of conservation of energy shows that, when only conservative forces act on an object, the object's total mechanical energy—the sum of its kinetic and potential energy—is unchanged. Conservation of energy provides a "shortcut" to understanding motion, allowing us to predict aspects of complicated motion without getting into intractable mathematical details. Sometimes we're concerned not so much with energy as with the rate at which we do work, consume energy, or transform energy from one kind to another—power describes this rate.

- We're going to explore further the implications of the law of conservation of energy, which is the idea that we can transform energy from potential energy to kinetic energy and vice versa in systems where only conservative forces act so that the total mechanical energy—the sum of the kinetic energy and the potential energy—will be conserved.

- One of the biggest reasons for understanding conservation of energy is it allows us to shortcut the details of problems involving motion that would be difficult to do using Newton's second law.

- It's pretty straightforward to solve problems involving motion with constant acceleration using Newton's second law, but forces aren't always constant.

- One way to solve problems involving nonconstant forces is to use calculus, but sometimes that's not even possible mathematically.

- Let's look at an example that uses the law of conservation of energy. There are 2 skiers who each start from rest and ski down frictionless slopes. They're going to drop through a vertical distance *h*.

- The first skier is on a perfectly smooth slope that is constant. That skier is going to be accelerating downward with a constant acceleration, so we could easily answer the question: How fast is the skier going at the bottom?

- The second skier is on a slope whose angle changes. This skier is going to start out with a relatively small acceleration, go over a hump during which the acceleration is going to increase, and then the acceleration is going to decrease toward the bottom.

- Conservation of energy is going to cut through the details so that these 2 problems become essentially the same. Let's start by setting the zero potential energy at the bottom of the slope.

- Then, the initial potential energy of either skier is U_0, which equals *mgh* because they are starting out at height *h,* and the bottom is where the potential energy is assumed to be zero.

- The skiers start at rest, so their initial kinetic energies are $K_0 = 0$. Therefore, their total energy is $K_0 + U_0 = mgh$.

- At the bottom, $U = 0$, so their energy is all kinetic ($K = (1/2)mv^2$). We want to know how fast they're going at the bottom, which is *v*.

- Using the statement of conservation of energy, we know that their initial energy is equal to their final energy ($mgh = (1/2)mv^2$).

- The mass cancels because the equation involves gravity. Solving for *v*, we get $v = \sqrt{2gh}$.

- Regardless of how complicated the slope was, as long as there were no losses due to friction, the answer would be the same for any skier: They would reach the bottom with the same speed, $v = \sqrt{2gh}$.

- Energy, a conserved substance in the universe, is related to **power**, which measures the rate at which you do work, use energy, generate energy, and transform energy. It's a very common misconception to confuse energy and power.

A horsepower, a unit used to measure the power of car engines, is equivalent to 746 watts.

- We know that energy is measured in joules, but it's also measured in calories, kilowatt hours, British thermal units, and many other forms.

- The fundamental SI unit for power is the **watt** (W), which is defined as 1 joule per second. A horsepower, another power unit, is equivalent to 746 watts.

- What's the minimum power needed to bicycle up a 10% (6°) slope at 9 mph (4 m/s) if the total mass of you and your bicycle is 80 kg?

- We know that the work it takes to raise a mass m a height h is $W = mgh$. However, we need the rate of doing work, which is power (P).

- Power is going to be the rate at which we're raising the mass up in height times the rate of change of height over the time involved in making that change ($P = mg(\Delta h / \Delta t)$).

- The vertical speed is $\Delta h / \Delta t$, which we'll call $v_{vertical}$, the rate at which the bicycle is going up vertically. $P = mgv_{vertical}$.

- Using trigonometry, the sine of 6° is about 0.1, so the vertical speed is therefore 0.1 of the speed along the slope because 4 m/s is the hypotenuse of the triangle with the 6° angle. So, $v_{vertical}$ is 0.4 m/s.

- Let's plug in the numbers.

$$P = mgv_{\text{vertical}} = (80 \text{ kg})(9.8 \text{ N/kg})(0.4 \text{ m/s}) = 313 \text{ W}$$

- As the cyclist, you're putting out energy at the rate of 313 joules every second, and we're not even counting friction or air resistance. This gives us an idea of how much power our bodies can produce.

Important Terms

power: The rate of producing or expending energy.

watt (W): In the International System of Units, the rate of energy conversion equivalent to 1 joule per second. It is named for Scottish engineer James Watt.

Suggested Reading

Rex and Wolfson, *ECP*, chap 5.5–5.6.

Wolfson, *EUP*, chap 7.2–7.4.

Questions to Consider

1. Railroad cars are often sorted by humping—a procedure where an engine pushes the car to the top of a small hump and then lets gravity take it down through a switchyard and onto the appropriate track. At the end of each track is a spring bumper to prevent damage to the cars. Describe the interchanges between different forms of energy that take place from the moment the car is released from rest at the top of the hump until it's against the spring bumper at the point of the spring's maximum compression.

2. A new wind farm is proposed for your community, and the developer claims it will produce 10 megawatts of energy each hour. What's wrong with this statement?

Gravity
Lecture 13

Newton used his gravitational theory to explain the orbital motion of the planets and even foresaw the possibility of artificial satellites. Newton developed calculus to solve the problem of planetary motion, showing that the orbits of planets must be ellipses. The special case of circular orbits doesn't require calculus and yields insights into orbital energy and spacecraft maneuvers. Consideration of gravitational energy also shows that it takes a finite speed—the so-called escape speed—for an object to break its bond with Earth or any other gravitating body.

- Gravity is one of the fundamental forces of nature, but the early views of gravity weren't really about gravity at all—they were about the structure of the universe.

- The views of Aristotle and Ptolemy dominated for thousands of years and depicted a **geocentric** universe: a universe in which Earth was at the center and the Sun and other planets traveled around Earth.

- Nicolaus Copernicus (1473–1543) introduced the idea of the **heliocentric** universe, in which the Sun was at the center and the planets traveled around it. He persisted in the idea that all orbits were circular.

- As an assistant to the astronomer Tycho Brahe, Johannes Kepler (1571–1630) determined that the orbits of the planets were not circles but **ellipses**—special curves with 2 points called foci, which are like the center of a circle spread out into 2 points.

- Kepler studied elliptical orbits and came up with a number of laws that describe the orbits as well as how fast the planets move—more rapidly near the Sun and more slowly farther out.

- Around 1610, Kepler developed relationships between the period and the size of the orbit in what are known as Kepler's laws, which describe the planetary orbits quite accurately but didn't explain why they were that way.

- Galileo (1564–1642) was the first to train a telescope on the sky and made a number of observations that supported both the Copernican and Keplerian view that Earth was not at the center of the universe.

Nicolaus Copernicus (1473–1543) introduced the idea of a heliocentric universe.

- Galileo worked toward the realization that what happens on Earth and what happens in the universe is not a different realm of physical reality—similar considerations apply to both.

- Isaac Newton laid the foundations for calculus and his study of optics and gravity. In Newton's formulation of a law of gravity, he realized that gravity extended from what we recognize on Earth to a force that operated throughout the universe—an idea he called **universal gravitation.**

- Newton developed a formula that describes universal gravitation between any 2 masses in the universe in which the force of gravity is proportional to the 2 masses and inversely proportional to the distance between them.

- The force of gravity F is the force between 2 masses, m_1 and m_2, and r^2 is the square of the center-to-center distance between the masses. The constant of universal gravitation is described by G:
$$F = \frac{Gm_1m_2}{r^2}.$$

- Newton's universal gravitation obeys Newton's third law, and it keeps the Sun and the Solar System pulled toward the center of the galaxy.

- Elliptical orbits are complicated because the distance and force are always changing. Therefore, Newton invented calculus, which he used to show that his equation of universal gravitation led to planetary orbits that were just as Kepler described.

- Newton also realized that there was the possibility that human beings could put things into orbit—that we could launch satellites that would be able to orbit Earth.

- A **geosynchronous orbit**—a low-Earth orbit in which the orbital period is 24 hours—is the most valuable place in the real estate of space that is near Earth.

- Through mathematical calculations, we find that a geosynchronous orbit exists at an altitude of about 22,000 miles. A satellite in geosynchronous orbit remains fixed over a point on the equator.

- In Lecture 11, we found that the potential energy is the area under the force-versus-distance curve. In this case, the force of gravity falls off as the inverse square of distance.

- Gravity is an attractive force, which means that as an object moves closer to a gravitating mass like Earth, work is done on it.

- Because negative work is done on the object, the associated potential energy is negative—the area under the curve.

- As an object moves closer to the gravitating object, a substantial amount of potential energy is involved, and as it moves away (because the force is falling off so rapidly), less potential energy is involved.

- Newton's calculus shows that if you take zero potential energy to be way off in infinity and then you bring an object closer to a gravitating center, the potential energy goes down, meaning that it would take work to move that object farther away from Earth.

- The calculus shows that gravitational potential energy depends inversely on r (whereas the force of gravity depended inversely on r^2): $U = -\dfrac{GMm}{r}$. Like the formula for universal gravitation, this formula uses the same gravitational constant, mass of the object doing the gravitating, and mass of the object moving near it.

- Using the gravitational potential energy associated with 2 masses, M and m, a distance r apart, the zero of gravitational energy is at infinity when they're very far apart.

- The minus sign tells us that the gravitational potential energy is negative; therefore, we're dealing with an attractive force. In other words, it would take work to separate 2 objects that are connected gravitationally that are undergoing this attractive gravitational interaction.

- The total energy of an object m is the kinetic energy plus the potential energy of an object moving in the vicinity of a large mass M that's exerting a gravitational force on it: $E = \dfrac{1}{2}mv^2 - \dfrac{GMm}{r}$.

- If the total energy of an object moving in the vicinity of another mass is less than zero, it is in a bound (elliptical) orbit; if it's greater than zero, it's in an open (hyperbolic) orbit and will never return; and if it equals zero, the object has just enough energy to escape the mass (parabolic).

- In the borderline case when total energy equals zero, the dividing line defines a special speed—the escape speed—at which you would have to launch something from a gravitating object for it to be able to escape that object forever.

Important Terms

ellipse: A particular kind of stretched circle; the path of planets in orbit.

geocentric: The belief that the Sun and the entire universe rotates around Earth.

geosynchronous orbit: An equatorial orbit at an altitude of about 22,000 miles, where the orbital period is 24 hours. A satellite in such an orbit remains fixed over a point on the equator.

heliocentric: The belief that Earth and the rest of the Solar System revolve around the Sun.

universal gravitation: The concept, originated by Newton, that every piece of matter in the universe attracts every other piece.

Suggested Reading

Rex and Wolfson, *ECP*, chap 9.

Wolfson, *EUP*, chap 8.

Questions to Consider

1. Describe some of Galileo's astronomical observations that helped to break the dichotomy between the terrestrial and celestial realms.

2. In what sense are a falling apple and the Moon undergoing the same type of motion?

3. Why are orbits at 22,000 miles altitude considered among the most valuable real estate in space?

4. True or false: What goes up must come down. Explain your answer using the concept of escape speed.

Systems of Particles
Lecture 14

R eal objects consist of many particles. In principle, studying many-particle systems is complicated by the need to consider all the forces acting among the particles, as well as any forces applied from outside the system. But Newton's laws enable us to describe the overall motion of a complex system as though it consisted of a single particle located at a special point called the center of mass. Newton's third law, which describes the conservation of momentum, lets us analyze intense interactions called collisions.

- So far, we've been treating objects as simple entities whose only important attribute is their mass. Real objects, however, are more complicated than that—they're extended in space and often have parts that can move in relation to each other.

- We can use Newton's laws to describe the motion of complex systems in terms of a few fundamental properties of those systems.

- Objects have complicated motions, but for each object there exists one point that follows a very simple path in which the vertical position as a function of the horizontal position is a parabola—the path characteristic of projectile motion under the influence of Earth's gravity.

- Let's look at a paradigm of a 2-particle system. The 2 particles, or 2 masses, m_1 and m_2, are somehow connected to form a system. Suppose you apply some external force to that system and you want to understand the system's response to that force.

- What are all the forces that are acting on that system? That external force you're applying is not necessarily the only force acting; typically, the individual particles of a system are also interacting with each other (otherwise, the system would fall apart).

- There are interaction forces, typically electrical forces, that hold together the particles of a system. However, those forces are internal to the system and are therefore part of the system.

- When considering the entire system—not m_1 or m_2 individually— the 2 interaction forces cancel in pairs by Newton's third law. Therefore, the net force on the system is due to external forces only.

- The system responds to an external force as though it were a single particle located somewhere called its **center of mass**.

- Using a coordinate system, if m_1 is at position x_1 and m_2 is at position x_2, then the position of the center of mass in a 1-dimensional situation is given by a weighted average of the positions of the individual particles (in terms of their individual masses).

- The numerator of this expression is m_1x_1—the position of the first mass weighted by its mass—plus m_2x_2, which is then divided by the total mass of the system: $x_{cm} = \dfrac{m_1x_1 + m_2x_2}{m_1 + m_2}$.

- If the particles were arrayed in multiple dimensions, you could calculate a y_{cm} and a z_{cm} for the other 2 coordinates in 3-dimensional space.

- Newton's quantity of motion, momentum **p**, is mass times velocity, **p** = m**v**. If you have many particles in a system, the momentum of the entire system is simply the sum of the momentum of the individual particles.

- We will use capital letters for quantities that pertain to an entire system. So, *M* will be the mass of a system.

- The momentum is $m_1\mathbf{v}_1$, the product of the mass and velocity of the first particle making up the system, plus $m_2\mathbf{v}_2$, the second particle, plus as many particles as there are, which will be designated as *K*: $\mathbf{p} = m_1\mathbf{v}_1 + m_2\mathbf{v}_2 + K$.

- The total momentum of the system is the total mass of the system times a quantity $\Delta\mathbf{R}$ (a change in a position) divided by Δt, which is the change in the position of the center of mass. A change in position with respect to time is a velocity, so that expression can be replaced by $\mathbf{v}$, the velocity of the center of mass:

$$\mathrm{p} = M\,\frac{\Delta\mathbf{R}}{\Delta t} = M\mathbf{v}.$$

- In simpler terms, the momentum of a system is the mass of the entire system multiplied by the velocity of the center of mass.

- In addition to the law of conservation of energy discussed in Lecture 11, we now discuss the idea of **conservation of momentum**, which transcends classical physics and plays an essential role, for example, in quantum physics.

- Using Newton's second and third laws, we can derive an equation describing the effect on the momentum of a system of the net external force of a system: $\mathbf{F}_{net\ external} = \dfrac{\Delta\mathbf{p}}{\Delta t}$, which is the rate of change of momentum.

- This equation states that the momentum of the entire system will change if there's a net external force; consequently, in the absence of any external forces, a system's momentum does not change.

- Internal forces cannot do anything to the momentum of the system; they may reconfigure the system, but that cannot change the momentum of the entire system because all the forces involved are internal to the system.

- A commonplace occurrence in both the natural and technological world is a **collision**, an event in which a brief (meaning that the time involved is very short compared with, for example, the overall trajectories) and intense interaction occurs between 2 or more objects.

- By treating the colliding particles as a single system, we can learn quite a bit about collisions. The interaction between the 2 or more particles is so intense that we can ignore external forces; therefore, the total momentum of a system is essentially unchanged. Momentum is conserved in a collision.

- There are 2 extremes when dealing with the types of collisions in which momentum is conserved, but energy may or may not be. **Inelastic collision** is a type of collision in which kinetic energy is not conserved, and **elastic collision** is a type in which kinetic energy is conserved.

- If a block of mass m_1 strikes mass m_2 with some speed and velocity $\mathbf{v}_1$ and the collision is totally inelastic, they stick together and combine to make a single object that moves with speed $\mathbf{v}$, which is a little less than the initial speed.

- The initial momentum of a whole system is the momentum only of the incoming mass because it is the only one that is moving: $m_1\mathbf{v}_1$. The final momentum is the momentum of both masses that are moving with the same velocity: $(m_1 + m_2)\mathbf{v}$.

ACS Science & Engineering Team, NASA.

When studying a collision between 2 objects—in this case, a collision of galaxies that took millions of years to occur—we can learn a great deal by treating the colliding particles as a single system.

- Conservation of momentum says the initial momentum is equal to the final momentum, so if we solve for the final velocity $\mathbf{v}$, we get:
$$\mathbf{v} = \frac{m_1}{m_1 + m_2}\mathbf{v}_1.$$

- Elastic collisions are more difficult because both energy and momentum are conserved, and they become even more complicated if the collision is taking place in 2 dimensions with objects moving at different angles. Therefore, we're going to restrict the following equations to the special case of 1-dimensional collisions in which 1 object is initially at rest. As a result, we will not use vectors.

- Based on the conservation of momentum, $m_1 v_{1i} = m_1 v_{1f} + m_2 v_{2f}$. Using the law of conservation of energy, $\frac{1}{2}m_1 v_{1i}^2 = \frac{1}{2}m_1 v_{1f}^2 + \frac{1}{2}m_2 v_{2f}^2$.

- Algebraically, we can solve the 2 equations with 2 unknowns for the final velocities v_{1f} and v_{2f}: $v_{1f} = \dfrac{m_1 - m_2}{m_1 + m_2}v_{1i}$ and $v_{2f} = \dfrac{2m_1}{m_1 + m_2}v_{1i}$.

- If the first mass, the incoming mass, is much smaller than the other mass, you can neglect m_1 in the equation, which then equates to a negative velocity. Therefore, if a very light object hits a massive object, the light object bounces back with the velocity it came in with.

- If m_1 is much larger than m_2, then the velocity of the incoming object is essentially unchanged, and the struck object goes off with twice the velocity of the initial object.

- If m_1 and m_2 are equal, there's a complete transfer of energy the instant the first object stops, and the struck object moves on.

Important Terms

center of mass: An average position of matter in an object; the effective point where gravity (or external force) acts.

collision: An intense interaction between objects that lasts a short time and involves very large forces.

conservation of momentum: The situation that exists when the momentum remains unchanged during an interaction.

elastic collision: A collision in which energy is conserved.

inelastic collision: A collision in which energy is not conserved.

Suggested Reading

Rex and Wolfson, *ECP*, chap 6.

Wolfson, *EUP*, chap 9.

Questions to Consider

1. How is it possible for a high jumper to clear the bar when his center of mass doesn't?

2. Why doesn't a rocket need a medium, like air, to push against?

3. A collision between galaxies may take hundreds of millions of years. How is this consistent with the definition of a collision as a "brief" interaction between objects?

Rotational Motion
Lecture 15

R otational motion can be characterized in terms of quantities analogous to those describing ordinary motion: position, velocity, acceleration—which become angular position, angular velocity, and angular acceleration; force, which becomes torque; and mass, which becomes rotational inertia. Together, rotational analogs of acceleration, force, and mass obey a law analogous to Newton's second law. This, in turn, leads to the concept of angular momentum and the all-important conservation of angular momentum, which explains some surprising and seemingly counterintuitive phenomena involving rotating objects.

- In the case of **rotational motion**, motion about some axis, we seek to develop quantities that are analogous to the quantities we already understand for straight-line motion—quantities that are analogs of position, velocity, and acceleration.

- Rotational analogs of position (θ) and change in position ($\Delta\theta$) are displacement of velocity and of acceleration.

- **Angular displacement**, $\Delta\theta$, can be measured in revolutions, degrees, or radians. A **radian** is the natural measure of angle and, also, the official SI unit of angle; it is the ratio of the arc length to the radius on a circle or circular arc.

- Based on the equation for velocity, **angular velocity** becomes $\Delta\theta/\Delta t$, which we will denote as the symbol ω, the Greek letter omega.

- Angular velocity can be measured in revolutions per minute, degrees per second, or radians per second.

76

- Based on the equation for acceleration, **angular acceleration** (α) becomes the rate of change of angular velocity, $\Delta\omega/\Delta t$, which can be measured in revolutions per minute per second (rpm/s), degrees per second squared, or radians per second squared (s^{-2}).

- A radian is a dimensionless quantity because it's the ratio of one length to another, so we can say that radians per second is also just seconds to the -1 power, or inverse seconds (s^{-1}).

- There are 2π radians in a full circle because the circumference of a circle is 2π times its radius, and a radian is the ratio of arc length to radius. Therefore, 1 radian is 360° divided by 2π, which equates to about 57.3°.

- The linear velocity of a point that is located some distance r from the center of a rotating object all rotating with angular velocity ω is simply ωr.

- The analog of mass is **rotational inertia**, I, which is the resistance to changes in rotational motion.

- The rotational analog of force is called **torque** (τ), which is measured by the strength of the force and the perpendicular distance from the pivot point to the point where the force is applied.

- The rotational analog of Newton's second law is torque (τ) equals rotational inertia (I) times angular acceleration (α). ($F = ma$ now becomes $\tau = I\alpha$.)

- Linear kinetic energy ($K = 1/2mv^2$) is sometimes called **translational motion**, which means moving from place to place. Rotational kinetic energy, K_{rot}, is 1/2 times the analog of mass times the square of the analog of velocity ($K_{rot} = 1/2I\omega^2$).

- The analog of momentum (**p** = *m***v**) is angular momentum, which we will designate as *L*, equals rotational inertia (*I*) times angular velocity (*ω*).

- Rolling motion is a combination, or sum, of translational motion (motion of the center of mass) and rotational motion (motion about the center of mass).

- The center of mass is moving at v_{cm}, the velocity of the center of mass, and the rolling motion is the motion of the individual points on a rolling object.

- If we look at the 2 velocity vectors from the translational motion and the rotational motion, they add up to zero, which means that the bottom of the rolling object is at rest—but only for an instant—in true tolling motion.

- There are 2 objects—a solid disc and a ring—on a ramp. They have equal mass and equal radii. In a race down the ramp, the solid disc will reach the bottom first.

- The solid disc has lower rotational inertia because more of its mass is concentrated toward the center, making it easier to change its rotational motion; the ring has all its mass concentrated at its outermost edge, making it harder to change its rotational motion.
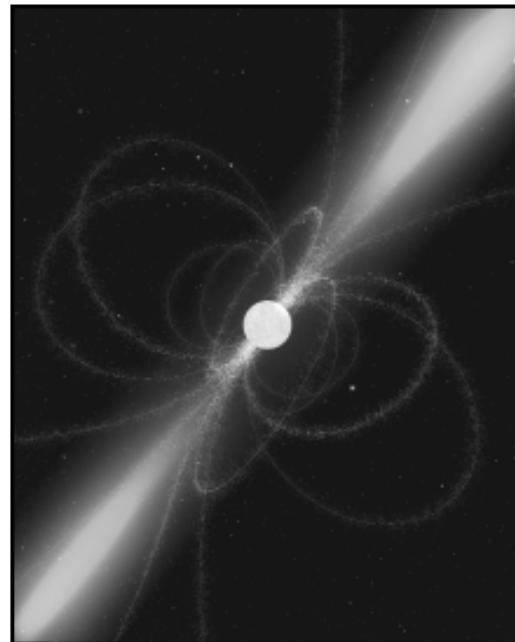
A pulsar is a rapidly spinning, extremely dense neutron star that releases charged particles into an intense magnetic field that surrounds the star and rotates along with it.

Lecture 15: Rotational Motion

- When these 2 objects start down the ramp, it is harder to accelerate the ring—it is harder to give it the rotational motion it needs to get down the ramp—so the solid disc wins.

- Mathematically, these 2 objects begin at some height h on the ramp and with potential energy $U = mgh$. They have no kinetic energy ($K = 0$) at the top of the ramp because they both begin at rest.

- At the bottom, both objects have potential energy $U = 0$ and kinetic energy $K = K_{trans} + K_{rot} = (1/2)mv^2 + (1/2)I\omega^2$. Their total energy consists of a combination of translational kinetic energy, $1/2mv^2$, plus rotational kinetic energy, $(1/2)I\omega^2$.

- Conservation of energy tells us $mgh = (1/2)mv^2 + (1/2)I\omega^2$. In rolling motion, $v = \omega r$, which means that $\omega = v/r$.

- Rotational inertias for the 2 objects are $I_{solid} = (1/2)mr^2$ and $I_{hollow} = mr^2$.

- Substituting v/r in for ω, the conservation of energy equation becomes $mgh = (1/2)mv^2 + (1/2)(amr^2)(v/r)^2 = (1/2)mv^2(1 + a)$.

- Solving for velocity, the solid ends up with a velocity that is 82% of $\sqrt{2gh}$.

- For the hollow one, it's only $0.71\sqrt{2gh}$. Consequently, the solid disc wins, which is indeed what happens.

- Newton's second law written in terms of momentum is $\mathbf{F} = \Delta\mathbf{p}/\Delta t$. Therefore, the rotational analog becomes torque equals the change in angular momentum per time: $\tau = \Delta L/\Delta t$.

- We know that when a system is subject to no external forces, its momentum is conserved. Therefore, if there are no external torques acting on a system, the angular momentum of the system cannot change, or must be conserved.

- Just as Newton's law is about systems changing their motion in the direction of the force—not moving in the direction of the force—so Newton's law analog is about rotating systems changing their angular momentum in the direction of the torque that's acting on them.

Important Terms

angular acceleration: The rate of change of angular velocity.

angular displacement: Rotational analog of change of position.

angular velocity: A measure of the rotation rate of a rotating object.

radian: The natural measure of angle and, also, the official SI unit of angle; it is the ratio of the arc length to the radius on a circle or circular arc.

rotational inertia: A measure of an object's resistance to change in rotational motion.

rotational motion: Motion about some axis.

torque: The rotational analog of force; torque depends on force and where that force is applied.

translational motion: Moving from place to place.

Suggested Reading

Rex and Wolfson, *ECP*, chap 8.1–8.6, 8.8–8.9.

Wolfson, *EUP*, chap 10–11.

Lecture 15: Rotational Motion

Questions to Consider

1. Explain why a solid sphere has lower rotational inertia than a hollow sphere of the same mass and radius.

2. It's sometimes said that you can distinguish a hard-boiled egg from an uncooked one by rolling the 2 down a ramp. How does this work? Which egg should reach the bottom first?

3. A spinning ice skater pulls in her arms, and her spin rate increases. Has her angular momentum increased? What about her rotational energy?

Keeping Still
Lecture 16

S kyscrapers, bridges, and other structures are in static equilibrium—a state characterized by having no tendency to accelerate or rotate. Thus, objects in static equilibrium must be subject to zero net force and zero net torque. It's the job of engineers to design structures so these conditions are met, thus ensuring that structures will be in static equilibrium. But even that's not enough: Equilibrium should be stable, meaning that small disturbances that disrupt the system's state of equilibrium will result in the system returning to equilibrium.

- **Static equilibrium** is a special case in which objects are neither accelerating in the linear fashion, nor undergoing angular acceleration, or even rotating. The requirements are that the system has zero net force and zero net torque.

- Because torque depends not only on the force you apply but on how far it is from the pivot point, the problem with static equilibrium is that there isn't any rotation, so there isn't an obvious choice of pivot point.

- As long as the net force of the object is zero (a condition for static equilibrium) and the torque about any point is zero, we can prove that the torque about every point is also zero.

- Every object has a point called the **center of gravity**, and it's as if all the mass of that object were located at this point, at least for the purposes of the torque that gravity exerts on the object.

- For the purposes of torque, the center of gravity is the place where gravity appears to be acting. For the purposes of Newton's second law, the center of mass is the point at which the mass seems to be concentrated.

82

- When gravity is uniform, the center of gravity corresponds with the center of mass.

- When the only torque is due to gravity, then the center of gravity must lie directly above the pivot point. If the center of gravity is a little bit off-center, there will be a torque.

- When there is only a single point of contact with the ground or with a pivot point, then the center of gravity must be on a vertical line with that point. Otherwise, the system would not be in static equilibrium.

- The coincidence of the center of gravity and the center of mass is going to depend crucially on gravity being uniform, but it's approximately uniform near Earth's surface.

Because the force of the ground balances the force of gravity on the Leaning Tower of Pisa, there is no net force on it and, consequently, it does not topple.

- The torque due to gravity on any object can be calculated as if the object were a single object located at its center of mass relative to the pivot point. We then multiply that position of the center of mass relative to the pivot by the gravitational force on the entire system:

$$\tau = \left(\frac{L_1 m_1 + L_2 m_2 + \dots}{M} \right) Mg.$$

- In Figure 16.1, a ladder is leaning against a wall, and you want to know at what angle can you prop it and still safely climb to some height before the ladder begins to slip out at the bottom?

Figure 16.1
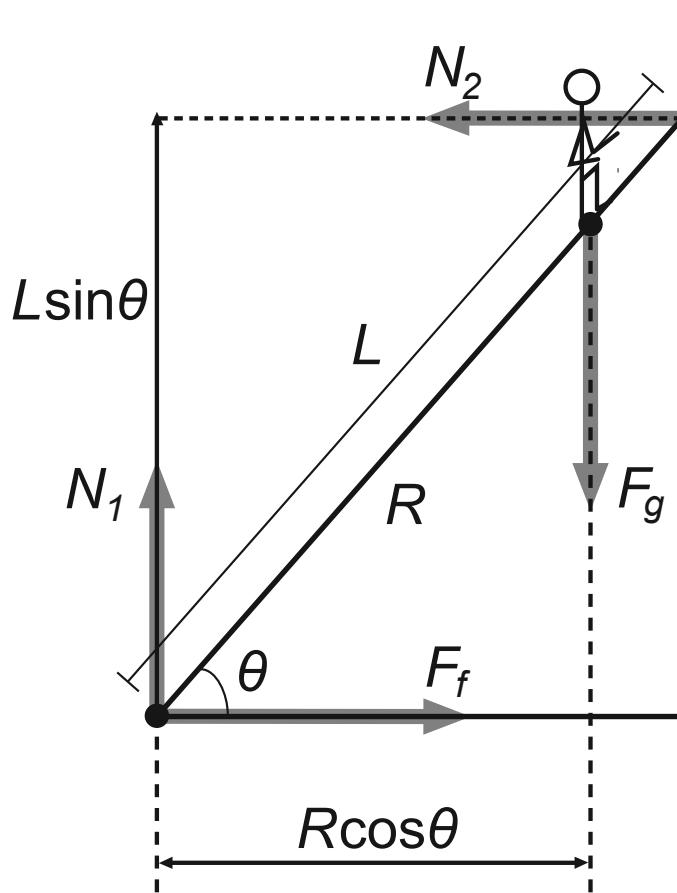
- To begin with, the ladder has negligible mass, and we can assume that the wall is frictionless.

- There is friction at the ground, and the coefficient of friction is μ. It happens to be the coefficient of static friction because we want the ladder to be at rest with respect to the ground and not slip.

- In mathematical terms, what distance R can you, mass m, climb up the ladder before the ladder slips?

- As a reminder, friction is proportional to the normal force between the 2 surfaces that are touching. The force of friction is μ, the coefficient of friction, times the normal force.

- We are free to choose the pivot point anywhere, so we choose the pivot at the bottom to simplify the math.

- First, we need to identify the forces acting on this system, which is the ladder plus the person: The force of gravity is acting on the person downward, and there's no force of gravity on the ladder because the ladder has negligible mass.

- There is a normal force (N_1) from the ground pushing up vertically on the ladder because the ground is horizontal. There is also a normal force (N_2) from the wall pushing perpendicularly to the wall in the horizontal direction.

- Because the ladder is trying to slip out at the bottom, the frictional force is going to go the opposite way, F_f.

- The whole ladder has length L, and the angle is θ. The side opposite the ladder is $L\sin(\theta)$.

- One of the perpendicular distances that needs to be identified for finding torques is the distance from the pivot point; for N_2, this distance is $L\sin(\theta)$.

- The lever arm for the force of gravity is perpendicular distance from the pivot point to the line through that force; this distance R, a hypotenuse, has a cosine of the angle because this distance is the adjacent side to the angle.

- The vertical forces are the gravitational force downward and the normal force from the ground up. In equilibrium, the forces and the torques both have to sum to zero so that the ladder won't be accelerating in the vertical direction.

- In a vertical coordinate system, we can simply say that the magnitude of the upward normal force is equal to the magnitude of the downward gravitational force, or $N_1 = mg$.

- The horizontal forces are the force of friction and the normal force from the wall, which also have to balance. There's no acceleration in the horizontal direction, so the force of friction plus the normal force from the wall have to sum to zero.

- In a horizontal coordinate system, we can simply say that the magnitudes of the 2 horizontal forces are equal, or $N_2 = \mu N_1$.

- The torque due to the gravitational force is the force (mg) times the lever arm, $R\cos(\theta)$, which has to equal N_2 times its lever arm, $L\sin(\theta)$: $mgR\cos\theta = N_2 L \sin\theta$.

- Algebraically, we have 3 equations and 3 unknowns, which we can solve to get the equation $R = \mu L \dfrac{\sin\theta}{\cos\theta} = \mu L \tan\theta$.

- The bigger the coefficient of friction, the higher you can climb. The longer the length of the ladder, the higher you can climb, at least relative to that length. Finally, the bigger the tangent of θ—the steeper the angle—the higher you can climb.

- What's required for **stable equilibrium** is that a system must not only be in equilibrium—zero net force and zero net torque—but it must also be at a minimum in its potential-energy curve.

- Unstable equilibrium, if disturbed, is likely to end up in a stable equilibrium because it's likely to change the configuration it's in. **Metastable equilibrium** is an equilibrium that is neither fully stable nor fully unstable.

- Equilibrium becomes more complex when working in multiple dimensions; in 2 dimensions, a structure might be stable in one direction but unstable in another direction.

Important Terms

center of gravity: For the purposes of the torque that gravity exerts on the object, the point at which an object's mass acts as if all the object's mass were concentrated.

metastable equilibrium: An equilibrium that is neither fully stable nor fully unstable.

stable equilibrium: A system in equilibrium—zero net force and zero net torque—that must also be at a minimum in its potential-energy curve.

static equilibrium: A state in which an object is subject to zero net force (and zero net torque) and does not feel a large or increasing force if it is moved.

Suggested Reading

Rex and Wolfson, *ECP*, chap 8.7.

Wolfson, *EUP*, chap 12.

Questions to Consider

1. Under what conditions would an object's center of gravity and center of mass not coincide?

2. Highly polished surfaces can actually exhibit greater friction than less-polished surfaces. Why might this be the case?

3. A system has a potential-energy curve marked by hills and valleys, with slopes in between. Where will you find its stable equilibrium configurations?

Back and Forth—Oscillatory Motion
Lecture 17

Oscillations occur when a system that has been disturbed from stable equilibrium experiences a force that tends to restore the equilibrium; if it didn't, its equilibrium wouldn't be stable. When the restoring force is directly proportional to the displacement from equilibrium, the resulting oscillation is called simple harmonic motion (SHM). When a system is subjected to forces that vary at or near its natural oscillation frequency, resonance results in the buildup of large-scale oscillations, which can be damped by friction and related forces.

- This lecture on oscillatory motion is the first of Section 2, which applies to the ideas of Newtonian mechanics and more complex motions such as oscillations, waves, and fluids.

- Systems tend naturally to find themselves in a state of equilibrium, but if they're not, they quickly try to reach that state so that they do not remain unstable. If you disturb a system that's in its stable equilibrium state, it's going to eventually return to that equilibrium.

- Systems have inertia, though, so they tend to overshoot their equilibrium position, which leads them into a back-and-forth oscillatory motion—a universal phenomenon that occurs throughout the universe.

- A **period** (T) is the time for 1 complete oscillation cycle, and its SI unit is the second. The **frequency** is the number of oscillation cycles per second, and its SI unit is cycles per second, or hertz (Hz). The frequency and the period are inverses of each other.

- **Amplitude** is the maximum disturbance from equilibrium.

- From molecular systems to the astrophysical scale, the forces—torques, for example—that tend to restore the system to equilibrium are proportional to the placement, or how far it has moved from equilibrium.

- **Simple harmonic motion** is the motion that occurs when the restoring force or torque is directly proportional to displacement from equilibrium. This motion is characterized by a simple relationship between the position of the object undergoing the motion and the time.

- An important feature of simple harmonic motion that does not necessarily apply to more complicated oscillatory motions is that the frequency and period are independent of the amplitude.

- The spring is the paradigm of simple harmonic motion; pulsating stars, vibrating molecules, and swaying skyscrapers are all examples of systems that undergo simple harmonic motion.

- In a mass-spring system, Newton's second law tells us that $F = ma$, and a spring produces a force $-kx$. (The minus sign is there because if you stretch the spring a distance x, the force tends to pull back in the other direction.)

- This is a horizontal spring system that is oscillating back and forth. Of course, acceleration is the rate of change of velocity, and velocity is the rate of change of position.

- Newton's second law then becomes:

  $-kx$ = (mass)(rate of change of (rate of change of position)).

- Using calculus, we obtain the position of the object in a mass-spring system as a function of time (in which the constant A is the amplitude of the motion), $x = A\cos \omega t$, and the angular frequency (which is related to angular velocity and uses the same symbol), $\omega = \sqrt{k/m}$.

- For the equation $x = A\cos\omega t$, x is the position of the mass m as a function of time, and $x = 0$ is the position of the spring when it's unstretched, or in equilibrium.

- The maximum value x ever achieves is the amplitude, A, because x swings between positive and negative A. The angular frequency is ω, and t is time. We use cosine because we choose time to equal zero at the maximum displacement.

- For the equation $\omega = \sqrt{k/m}$, angular frequency has units of radians per second or s^{-1}; the frequency will be higher if the spring is stiffer (k is larger), and the frequency will be lower if m is larger.

- Simple harmonic motion in a mass-spring system involves energy: It has kinetic energy of the mass as it moves, and potential energy of the stretched spring.

- In simple harmonic motion, energy is interchanged continually between kinetic and potential energy in a mass-spring system. If there's no mechanism of energy loss—no friction, for example— this interchange of energy will continue basically forever.

- The potential energy of a mass-spring system is the potential energy stored in the spring when it's stretched or compressed a distance x from its equilibrium, as given by $U = 1/2kx^2$.

- Another way of characterizing simple harmonic motion is that the potential energy is parabolic, or quadratic, in the displacement—it depends on x^2.

- Stable equilibrium occurs at minimum points in potential-energy curves, and most curves approximate parabolas at their minimum points.

- A pendulum exhibits simple harmonic motion for small-amplitude swings, where the parabola approximates a circular arc; a pendulum with a large amplitude would demonstrate more complicated motion, and its period would be independent of its amplitude.

- The period of a pendulum is $T = 2\pi\sqrt{L/g}$. In this system, gravity plays the role of the spring—it's what provides the restoring force that pulls the pendulum back toward its equilibrium.

If you are driving and hit a pothole, the car's spring system allows the wheel and car to move relative to each other vertically, but this motion is damped out by shock absorbers.

- The element of gravity appears in the same place in this equation that the spring constant did in the equation for the period of the mass-spring system. However, in the mass-spring system, the mass is what provided the inertia (the bigger the mass, the slower the oscillations). In this equation, it's the length of the pendulum string that provides the inertia.

- For a pendulum with a 1-second period, we could solve the equation for L to determine the length of the pendulum string: $L = g\,T^2/4\pi^2 = (9.8)(1^2)/4\pi^2 = 0.25 \text{ m} = 10"$.

- The potential-energy curve for a pair of hydrogen atoms has a minimum where the 2 atoms are located when they form the stable H_2 molecule.

- The electrical forces, the repulsion of the nuclei, and the attraction of electrons within a pair of hydrogen atoms result in spring-like behavior—at least for small disturbances from equilibrium—and undergo simple harmonic oscillations.

- Simple harmonic motions tend to lose energy in a process called **damping**. If you want to keep an oscillating system oscillating, this process is a nuisance, but there are many cases in which it is beneficial that an oscillating system be damped—for example, the role of shock absorbers in a car accident.

- Oscillating systems tend to have a frequency at which they would like to oscillate, and if you happen to hit them with something at that frequency, the phenomenon is called **resonance**.

- Resonance is an important phenomenon when you design any kind of structure, but you would especially like to avoid resonance at frequencies that a structure is likely to be subjected to.

Important Terms

amplitude: The size of the disturbance that constitutes a wave.

damping: The process by which simple harmonic motions tend to lose energy.

frequency: The number of wave cycles per second that pass a fixed point in space.

period: The time interval between 2 successive wave crests; equivalently, the time for a complete wave cycle.

resonance: In weakly damped systems, this is the ability to cause large-amplitude oscillations with relatively small force.

simple harmonic motion (SHM): The motion that occurs when the restoring force or torque is directly proportional to displacement from equilibrium. This motion is characterized by a simple relationship between the position of the object undergoing the motion and the time.

Suggested Reading

Rex and Wolfson, *ECP*, chap 7.

Wolfson, *EUP*, chap 13.

Questions to Consider

1. How are frequency and period of oscillatory motion related?

2. What distinguishes simple harmonic motion from other forms of oscillatory motion?

3. Explain why simple harmonic motion is common throughout the physical universe.

Making Waves

Lecture 18

When oscillatory motion occurs in a continuous medium, oscillations couple to adjacent parts of the medium, resulting in a wave that propagates through the medium. Waves are described by the size of the motion (amplitude), the rate at which the motion repeats (frequency), and the distance between wave crests (wavelength). Unlike particles of matter, 2 waves can coexist at the same point. When they do, they interfere, their wave amplitudes adding at that point. Wave interference results in a host of useful and surprising phenomena.

- Waves are among the most important phenomena in physics and in your everyday life—you wouldn't hear without sound waves, and you wouldn't see without light waves.

- Waves are closely related to oscillations except that in a wave, an oscillation occurs in some part of the material medium, and the disturbance is propagated to nearby regions. As a result, a propagating disturbance, the wave, moves through the medium.

- A **wave** is a traveling disturbance that carries energy, but not matter. A wave may involve matter or disturb matter, but a wave is not matter.

- A wave that results from a disturbance at right angles, or perpendicular, to the medium in which the wave is propagating is called a **transverse wave**. An example of a transverse wave is a light wave.

- A longitudinal wave happens if a disturbance occurs in the direction of the medium. Sound waves are examples of longitudinal waves.

- The graph of a sinusoidal, or sine, function—like the one in Figure 18.1—shows the position of x as a function of the sine of x.

Figure 18.1

sin(x)

Position, x

-6   -4   -2   2   4   6   8   10   12
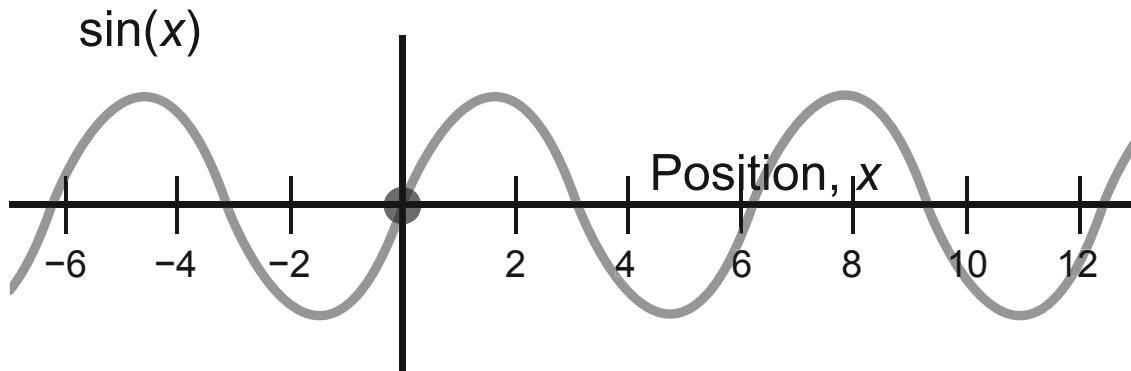
- If you wanted to plot $\sin(x - 2)$, move the whole sine wave to the right by 2 units on the x-axis. See Figure 18.2.

Figure 18.2

sin($x - 2$)

Position, x

-6   -4   -2   2   4   6   8   10   12
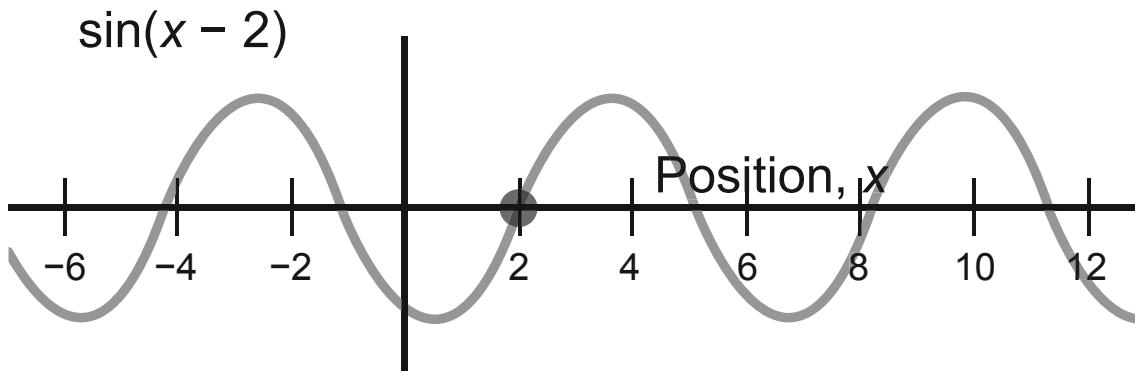
- A wave is described by a function of position and time, and the way position and time enter that function is in the form of position minus velocity times time ($\sin(x - vt)$). That's how we make a wave move. See Figure 18.3.

Figure 18.3

$\sin(x - vt)$

Position, x

-6   -4   -2   2   4   6   8   10   12
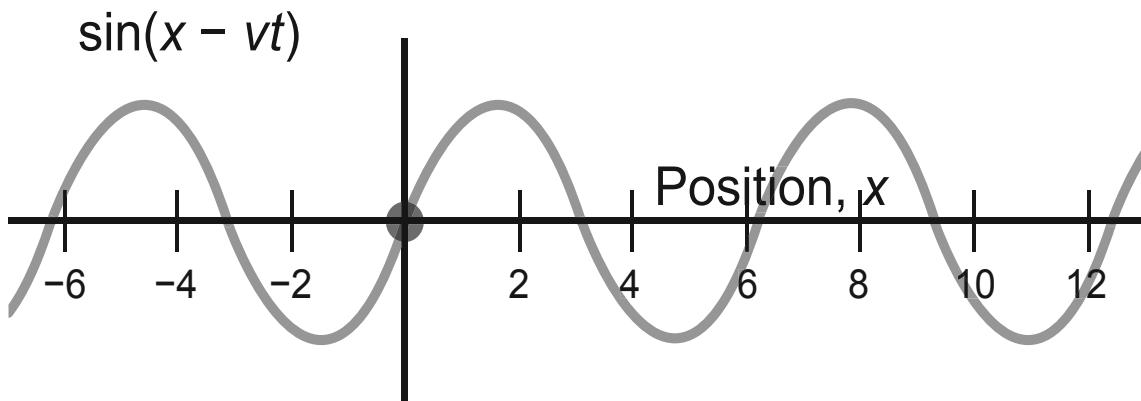
- At a fixed time, a sinusoidal curve gives the displacement y as a function of the position x from some amplitude A to the opposite of that amplitude $-A$. Between the crests of the wave is a number that is called the **wavelength**, λ, which is a distance.

- At a fixed position in space while time varies, there is still an oscillation with amplitude from A to $-A$, and there is a certain time it takes the wave to repeat itself, which is called the period of the wave, T.

- While the wavelength is the distance between wave crests, the wave period is the time between wave crests. The wave moves 1 wavelength λ in 1 period T.

- The wave speed is the wavelength divided by the period: $v = \lambda/T$. The wave frequency is 1 divided by the period, which is the same as it is for the oscillatory motion: $f = 1/T$. Therefore, $v = \lambda f$.

- Unlike matter, 2 waves can be in the same place at the same time—they can interfere with each other.

- Wave **interference** occurs when 2 waves are in the same place at the same time, and when they are, their amplitudes add if they're moving in the same direction or subtract if 1 wave is going down and 1 wave is going up.

- In quantum mechanics, we use wave interference to understand some very unusual phenomena: When dealing with sound waves, beats occur because of the interference of waves of nearly equal frequency, and they provide an audible indication of whether frequencies are synchronized.

- A pattern of wave interference in which 2 sources of waves are pulsing at the same frequency, causing them to send out wave crests that spread out farther and farther, is called **2-source interference**.

- Waves do not always have to be in movement. Sometimes, waves develop that oscillate in place and are called **standing waves**: waves that don't actually go anywhere but nevertheless have waves that go back and forth very rapidly, reflecting and interfering in a way that produces a kind of semi-fixed pattern.

The radar guns of police officers measure the Doppler effect of radar waves that bounce off your car.

- One of the most important wave phenomena is the **Doppler effect**, which occurs when there's relative motion between the source of waves and the observer of those waves. Under those conditions, there can be shifts in the frequency and wavelength of the waves, and those shifts can be used to measure, for example, relative speed.

- If the wave source is moving closer to an observer, he or she is going to measure a shorter wavelength because every time the source emits a wave crest, the source then moves closer to that observer.

- If the wave source is moving away from an observer, he or she is going to measure a longer wavelength—and a lower frequency as a result.

- The wave speed (v) is the wavelength divided by the wave period ($v = \lambda/T$).

- During 1 wave period T, the wave crests move 1 wavelength. The source, on the other hand, moving at speed u, moves a distance of speed times time, uT.

- In general, $\lambda' = \lambda\left(1 \mp \dfrac{u}{v}\right)$, where the negative sign is used when the wave source is approaching, and the positive sign is used when the wave source is receding.

- For objects coming toward you or going away from you—whether they are planets going around a distant star, pulsations of the Sun, or galaxies moving in the expanding universe—the Doppler effect provides a direct way to measure velocities of distant objects.

- A **shock wave** occurs when the source speed exceeds the wave speed, causing an intense pileup of wave fronts that cause a cone wave that is left behind from the moving source.

- The shock wave that occurs when aircraft are flying faster than sound is called a sonic boom—not just when they break the sound barrier, which is a common misconception, but the whole time they're moving faster than sound.

Important Terms

2-source interference: A pattern of wave interference in which 2 sources of waves are pulsing at the same frequency, which causes them to send out wave crests that spread out farther and farther.

Doppler effect: Named after a 19th-century Austrian physicist, this is the effect produced when the source of a wave and the observer of the wave are in relative motion. When the 2 are approaching each other, the wavelengths of the wave are compressed, leading to a higher pitch (in sound) or a bluer color (in light). When the 2 are receding, the distance between the wave crests is lengthened, leading to lower pitch or redder light.

interference: The process whereby 2 waves, occupying the same place at the same time, simply add to produce a composite disturbance. Interference may be constructive, in which the 2 waves reinforce to produce an enhanced composite wave, or destructive, in which case the composite wave is diminished.

shock wave: A very strong, abrupt wave produced when a wave source moves through a medium at a speed faster than the waves in that medium. An example is a sonic boom from a supersonic airplane.

standing wave: A wave that "stands" without propagating on a medium of fixed size.

transverse wave: A wave that results from a disturbance at right angles, or perpendicular, to the medium in which the wave is propagating.

wave: A traveling disturbance that carries energy but not matter. Waves may either be traveling (like a moving sound wave) or standing (like the vibrations of a wire with fixed ends).

wavelength: The distance between successive crests of a periodic wave.

Suggested Reading

Rex and Wolfson, *ECP*, chap 11.

Wolfson, *EUP*, chap 14.

Questions to Consider

1. Waves carry energy from the open ocean toward shore. Do they also move water shoreward?

2. Can 2 waves exist at the same point in space?

3. Police radar works by reflecting radio waves, which have short wavelengths, off a moving car and measuring the Doppler shift in the waves's frequency. The shift, in this case, is twice that given by the Doppler formulas developed in this lecture. Why?

Lecture 18: Making Waves

Fluid Statics—The Tip of the Iceberg
Lecture 19

A fluid can be at rest—in static equilibrium—only if the net force is zero. On Earth, that requires a balance between a fluid's pressure force and gravity. In equilibrium, the force of gravity is the weight of the overlying fluid layers; that's why water pressure increases with depth, and air pressure decreases with height in the atmosphere. The greater pressure with depth causes buoyancy, an upward pressure force that, for objects less dense than the fluid, results in a net upward force.

- **Fluids** are materials that are free to distort and change shape—such as liquids and gasses.

- In a liquid, the molecules are very close together, but they're free to move past each other, making them essentially incompressible In contrast, the molecules in a solid are locked in place in a regular crystalline structure.

- It's very difficult to change the density of a liquid, and we use the approximation that liquids have constant density.

- In a gas, on the other hand, the molecules are very distant and are moving around quite randomly. It's easy to change the volume of a gas.

- In both cases, the behavior of a fluid could be calculated by applying Newton's laws to all the individual molecules making up a fluid, but that would be a computationally impossible task. Therefore, we approximate that fluids are made up of continuous matter.

- **Density** is the mass per unit volume of the fluid. Its symbol is the Greek letter rho (ρ); its SI unit is kilograms per cubic meter (kg/m^3).

- **Pressure** is the force per unit area that a fluid exerts on its surroundings—on its container or on the adjacent fluid. Its symbol is *p,* and one of its SI units is newtons per meter squared; 1 newton per meter squared is called a pascal (Pa) after the mathematician Blaise Pascal.

- Other pressure units are pounds per square inch (psi), pounds per square inch gauge (psig)—which means how much pressure there is above atmospheric pressure called gauge pressure—and atmospheres (atm), where 1 atmosphere is defined as a standard value for the mean atmospheric pressure at sea level.

© iStockphoto/Thinkstock.

By turning on the hot-air jets, pilots can adjust the density of hot-air balloons, causing them to float at a constant altitude.

- An important thing about pressure is that it's exerted equally in all directions throughout a fluid. When a fluid is magnetized or electrically conducting, however, pressures are different in different directions.

- **Hydrostatic equilibrium** is the condition in which there is no net force on a fluid. In order to have a net force of zero, an upward force is required to balance the downward force of gravity, and that upward force has to be due to pressure.

- Pressure exerts a force per unit area, but the only time we experience a net force due to pressure is when there are pressure differences that give rise to forces.

- Figure 19.1 shows an element of fluid singled out from the rest that has an area A and a thickness Δh. We're going to designate h going downward as the depth of this fluid. There is a gravitational force downward on the fluid. There is a pressure force upward on the bottom of the fluid, which we will designate as $(p + \Delta p)A$.

Figure 19.1
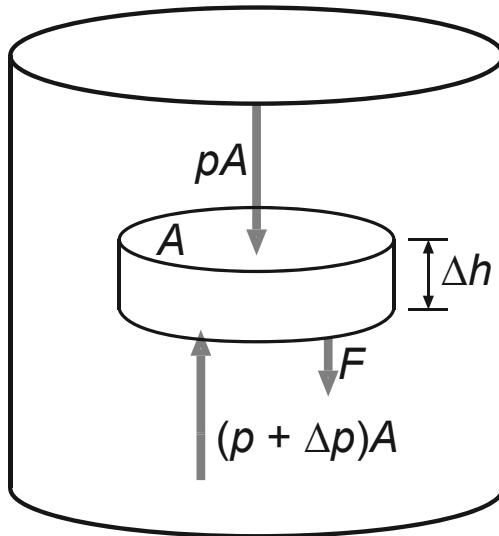
- The net force on this cylindrical fluid element is the difference between the downward force, pA, and the upward force, $(p + \Delta p)A$: $F_p = (p + \Delta p)A - pA = A\Delta p$.

- The upward pressure force is the area times the pressure difference across the distance Δh: $F_p = A\Delta p$.

- There is also a downward gravitational force mg. The mass of this particular fluid is the density of the fluid times the volume because density is mass per unit volume: $m = \rho A \Delta h$. Therefore, the downward gravitational force becomes $F_g = \rho A \Delta h g$.

- In equilibrium, the upward pressure force must equal the downward gravitational force: $\rho A \Delta h g = A \Delta p$. Algebraically, the A is going to cancel, and you end up with $\Delta p / \Delta h = \rho g$.

- In liquids where density is constant, the pressure increases uniformly with depth. In other words, the initial pressure is the pressure at the surface of the liquid, and then the pressure increases as we go downward: $p = p_{surface} + \rho gh$. This is the condition of hydrostatic equilibrium.

- Pressure ultimately results from a situation where there's gravity from the weight of the overlying fluid.

- Suppose a section of fluid has a height h, mass m, weight mg, and area A. The pressure is simply mg/A, but don't think of that as just a downward force because this is a fluid, which is free to deform. That pressure is exerted in all directions.

- For a liquid, the mass is the density times the volume (the height) times the area: $m = \rho hA$. Therefore, $p = \rho gh$ is the pressure at the bottom plus any extra pressure that was at the top.

- For gasses, density varies, but if the temperature is constant, calculus yields the fact that the pressure drops off exponentially from the surface: $p = p_{surface} e^{-h/h_0}$, where the exponent of e is the scale height. That's the approximate expression for Earth's atmospheric pressure.

- Another quality of static pressure is that pressure is transmitted equally. If you increase the pressure in a fluid—even by just a small amount—that increase is transmitted throughout the fluid.

- For a fluid that is in hydrostatic equilibrium, the upward pressure force—or buoyancy force (F_b)—balances the downward pressure force.

- If we were to replace the fluid with a solid object, the pressure force doesn't change, but the weight may. In equilibrium, the upward force (the buoyancy force on the water) and downward force (the gravitational force) are equal.

- If the gravitational force is less than the buoyancy force, the object is going to experience an upward force, causing it to float, which is true if the density of the object is less than the density of the fluid.

- If the gravitational force is greater, the object sinks, which is true if the density of the object is greater than the density of the fluid.

- **Archimedes's principle**, discovered in ancient times by the Greek mathematician Archimedes, says that the buoyancy force on a submerged object equals the weight of the displaced fluid.

- If you put an object under water, it pushes water out of the way—it displaces an amount of water equal to its volume—but the weight of that water is equal to the weight of a volume containing that same mass, so the buoyancy force becomes simply the weight of the displaced fluid.

- **Buoyancy** is a force that is caused by the pressure difference across a fluid, and this force remains acting on an object even after it has sunk. Objects that are under water weigh less—or, apparently weigh less.

- If an object is less dense than water, the buoyancy force will bring it up to the surface. **Neutral buoyancy** occurs when the object's density equals the fluid's density.

Important Terms

Archimedes's principle: Discovered in ancient times by the Greek mathematician Archimedes, this principle says that the buoyancy force on a submerged object equals the weight of the displaced fluid.

buoyancy: The upward force on an object that is less dense than the surrounding fluid, resulting from greater pressure at the bottom of the object.

density: The mass per unit volume of a fluid. Its symbol is the Greek letter rho, ρ, and its SI unit is kilograms per cubic meter.

fluids: Materials that are free to distort and change their shape, such as liquids and gasses.

hydrostatic equilibrium: The condition in which there is no net force on a fluid.

neutral buoyancy: The state of neither rising nor sinking that occurs for an object of the same density as the surrounding fluid.

pressure: The force per unit area.

Suggested Reading

Rex and Wolfson, *ECP*, chap 10.3–10.4.

Wolfson, *EUP*, chap 15.1–15.3.

Questions to Consider

1. Why would a water-based barometer be impractical? Roughly how high would it have to be?

2. A submerged object displaces a mass of water whose volume is equal to that of the object. What mass of water does a floating object displace?

3. With a hydraulic lift, a relatively small force can lift a massive vehicle. Are we getting something for nothing in this case? Discuss in terms of force and work.

4. Sea-level rise is a serious consequence of global warming. Will sea-level rise result from the melting of floating sea ice? Explain.

Lecture 19: Fluid Statics—The Tip of the Iceberg

Fluid Dynamics
Lecture 20

F luids are subject to 2 conservation laws: conservation of mass and conservation of energy. For a liquid, whose density cannot change significantly, conservation of mass requires high-flow velocities in which the liquid channel is constricted. Fluid pressure represents a form of internal energy, and therefore energy conservation requires low pressure where fluid velocity is high, and vice versa. This relationship between pressure and velocity results in many practical and sometimes counterintuitive phenomena collectively called the Bernoulli theorem. Fluids also exhibit a friction-like force called viscosity.

- **Fluid dynamics** is the study of the behavior of moving fluids. When fluids are moving, we have to address the flow velocity in addition to density and pressure.

- Just as density and pressure can be functions of position, so can the flow velocity. Furthermore, with fluid dynamics, all 3 of these quantities may vary with time even at a given position.

- **Steady flow** is a special case in which density, pressure, and velocity do not vary with time at a fixed position, although they may vary from position to position. The water isn't the same water at every given point, but every given point has a definite flow velocity.

- The mathematical treatment of unsteady flows, the phenomenon called turbulence, is an almost intractable mathematical problem that remains in an area of contemporary study in fluid dynamics, so we're only going to address steady flows.

- Moving fluids are represented by drawing a flow vector at every point in the fluid (or at representative points) or by drawing stream lines—continuous lines that are everywhere in the flow direction and that show the direction and the magnitude of the flow velocity.

- There are 2 conservation laws that describe flowing fluids: conservation of mass and conservation of energy.

- Preliminarily, a stream tube, or a flow tube, is a narrow region in a flow that is defined by adjacent or nearby stream lines. It is a tube that is small enough that fluid properties—pressure, velocity, density—don't vary significantly across the area bounded by the tube.

- In steady flow, fluid comes into the tube at the same rate that it leaves. We will assume that 2 flow elements have the same mass.

- The mass of a fluid element with density ρ, area A, and length L is the density times the volume, $\rho A L$.

- The fluid element crosses the area of the flow tube in time interval Δt with speed v: $L = v\Delta t$. Therefore, the mass of the fluid element is ρ times the area times the length: $m = \rho A v \Delta t$.

Wind turbines need to be located at a high altitude because viscosity slows down the air near the ground. Less energy can be extracted from slower-moving wind.

- The equation applies at both ends of the flow tube, and the masses are the same even though the density, area, and speed may vary. The time is also the same because it takes the same amount of time for that mass m to enter as for that mass m to leave.

Lecture 20: Fluid Dynamics

- In steady flow, the masses entering and leaving the flow tube in time Δt are the same, so $\rho_1 A_1 v_1 \Delta t = \rho_2 A_2 v_2 \Delta t$. Canceling the time element, $\rho_1 A_1 v_1 = \rho_2 A_2 v_2$.

- Equivalently, the quantity $\rho A v$ (in kilograms per second) is constant along the flow tube (not perpendicular to the flow), which means that mass is conserved in a steady flow. This equation is called the **continuity equation**.

- For liquids, the density (ρ) is a constant because the molecules in liquid are close together, which makes them difficult to compress. This property is also approximately true for gasses at subsonic speeds—at speeds that are much slower than the speed of sound.

- If we cancel the ρ because it's a constant, we are left with the volume flow rate Av along the flow tube with units in cubic meters per second.

- **Bernoulli's theorem** expresses the conservation of energy in a flowing fluid. We're going to restrict our study to **incompressible fluids**, meaning fluids for which the density is constant.

- A flow tube rises in height an amount Δy from one end to the other. There's a pressure p_1 at one end, an area A_1, fluids entering the tube with velocity v_1, and a fluid Δx_1 entering the tube. We want to know how much work is done on that fluid element as it moves—into the flow tube and out of the flow tube.

- We will use the work-energy theorem from Lecture 11, which says that the change in kinetic energy is equal to the net work done on an object—work is force times distance.

- The pressure at the inlet (p_1) times the area times the distance the fluid moves is $p_1 A_1 \Delta x_1$. At the exit, there's a pressure p_2 from the external fluid, which is pushing back, so it is doing negative work on the fluid: $-p_2 A_2 \Delta x_2$.

- As the fluid rises, gravity is pushing downward, so the fluid is working against gravity. Gravity, then, does negative work: $-mg$ times the height, Δy, or $y_2 - y_1$.

- There are 3 forces: incoming work from pressure, outgoing work from pressure, and the work done by gravity. The net work done on that fluid element as it goes through the tube is the sum of those 3 forces: $p_1 A \Delta x - p_2 A \Delta x - mg(y_2 - y_1)$.

- The work-energy theorem says that the net work done will be equal to the kinetic energy change, which is $\frac{1}{2}mv_2^2 - \frac{1}{2}mv_1^2$. Equating those and dividing by the volume $A\Delta x$ of the fluid in each case, we get $p_1 - p_2 - \rho g(y_2 - y_1) = \frac{1}{2}\rho v_2^2 - \frac{1}{2}\rho v_1^2$.

- Algebraically, $p_1 + \frac{1}{2}\rho v_1^2 + \rho g y_1 = p_2 + \frac{1}{2}\rho v_2^2 + \rho g y_2$. Equivalently, the quantity $p + \frac{1}{2}\rho v^2 + \rho g y$ is constant along the flow tube.

- The equation we just derived is Bernoulli's equation, which expresses the conservation of energy in a flowing fluid (restricted to incompressible fluids).

- The pressure (an internal density) plus the kinetic energy density plus the gravitational potential energy density add up to a constant, but individually they can change; you can change energy back and forth among different forms in this flowing fluid.

- Frequently, we deal with cases where there's no height change, and when that is the case, we can drop the $\rho g y$ because it's constant. Then, we are left with pressure plus kinetic energy density adding up to a constant, which is Bernoulli's equation: $p + \frac{1}{2}\rho v^2 = $ constant.

- The Bernoulli theorem states that in places where pressure is high in a flowing fluid, other places on the same stream tube where pressure is lower will have to have a higher velocity.

Lecture 20: Fluid Dynamics

- Bernoulli's theorem is only true as you move along a flow tube, and it leads us to assume that the presence of high pressure occurs with low speed, and vice versa.

- Sometimes fluid energy is not conserved. Fluid does have a kind of friction called viscosity, and it's particularly important when the fluid is flowing near a stationary material boundary—like the walls of a pipe.

- In these cases, instead of the flow being uniform across the pipe, it's zero at the edges where the fluid is literally attached to the boundary, and then it increases toward the middle of the pipe. The impact of this effect depends on the width of the tube in relation to the properties of the flowing fluid.

Important Terms

Bernoulli's theorem: A statement of energy conservation in a fluid, showing that the pressure is lowest where the flow speed is greatest and vice versa.

continuity equation: A statement of mass conservation in a steady flow, stating that the product of density, area, and speed (a quantity expressed in kilograms per second) is constant along the flow tube.

fluid dynamics: The study of the behavior of moving fluids.

incompressible fluids: Fluids for which the density is constant.

steady flow: A special case in which density, pressure, and velocity do not vary with time at a fixed position, although they may vary from position to position.

Suggested Reading

Rex and Wolfson, *ECP*, chap 10.5–10.6.

Wolfson, *EUP*, chap 15.4–15.6.

Questions to Consider

1. The Greek philosopher Heraclitus famously said, "You cannot step twice into the same river." Comment on the validity of Heraclitus' statement for steady and turbulent flows.

2. Why does blood pressure go down, instead of up, at the site of an obstruction in a blood vessel? Explain using the principles behind both the continuity equation and Bernoulli's equation.

Heat and Temperature
Lecture 21

T hermodynamics is the branch of physics that deals with heat, temperature, and related phenomena. Heat is a flow of energy that is driven by a temperature difference. One consequence of heat flow is to change an object's temperature. Specific heat is a property of materials that determines the heat needed for a given temperature change. When 2 objects at different temperatures are placed in contact, they come to thermal equilibrium, eventually reaching a common temperature.

- Section 3 deals with heat, which has a definition in physics that is different from what you might think it is. The topic of heat includes such subjects as thermodynamics and statistical mechanics.

- Heat is crucial to understanding energy flows throughout the universe—both naturally and technologically—from the astrophysical universe to the geophysical universe.

- Before 1800, heat was believed to be a fluid that flowed from hot objects to cooler objects. In the late 1700s, Benjamin Thompson studied the boring of canons and determined that heat was associated with doing mechanical work.

- In the 1840s, James Joule quantified the relationship between heat and energy—between calories (the measure of heat) and joules (the measure of mechanical energy), as it became known after his time.

- There are 2 objects in contact, which means they can exchange heat. If no **macroscopic properties** of them change—there are no measurable changes—then they are and were at the same temperature. If some properties change, then they were not at the same temperature, and they will eventually become the same temperature.

- You can choose any macroscopic property you want to define temperature: volume, length, electrical properties, or even properties of radiation.

- One macroscopic property is used as a standard way of measuring temperature: the pressure of a gas at a constant volume. As temperature increases, pressure increases.

- The temperature of water defines our temperature scales. At its **triple point**—the point at which liquid water and solid water, or ice and water vapor, can coexist—water has a unique temperature and pressure. In the conditions of the triple point, the temperature of water is exactly 273.16°C.

- At zero pressure, the temperature is the lowest possible temperature, −273°C, which defines **absolute zero**: the temperature at which the energy of a system is the minimum it can possibly be.

- The official scientific temperature scale of the SI system is the Kelvin scale, which is named after William Thomson, Baron Kelvin, and begins at absolute zero. The degree size in the Celsius scale is the same size as a kelvin, but the zero point is different—the zero in the Celsius scale is the melting point of ice instead of absolute zero, as in the Kelvin scale.

© Comstock Images/© Getty Images/Thinkstock.

If there's a power failure and your house drops to 5°C, it will take 3.5 hours to bring your house to room temperature if it has a heat capacity of 25 mJ/°C because ΔQ = CΔT.

- In the United States, we tend to use the Fahrenheit scale, in which ice melts at 32°F and water boils at 212°F. Absolute zero is about −460°F.

- Some of the most basic equations of thermodynamics require that absolute zero actually be zero on the scale, which is why the Kelvin scale is used.

- A temperature change of 1°C is the same as a temperature change of 1 K, which equals 1.8°F.

- **Heat** is energy being transferred from one object to another as a result of a temperature difference; heat is energy in transit—a flow of internal energy.

- There are several ways to change the internal energy of a system, and only one of them is a flow of heat. In other words, it may take a heat flow to change the internal energy of a system, but the change may be caused by mechanical energy, for example.

- **Heat capacity** is a measure of how much heat can flow into an object for a given temperature difference. The word "capacity" might sound like the object is holding heat, but it isn't.

- The most common result of heat flow is temperature change; for example, if heat flows into an object, the temperature increases.

- The temperature change ΔT is proportional to the amount of heat that flowed to a particular object, which is the heat capacity, C. The SI units for heat capacity are J/K and, equivalently, J/°C.

- A more fundamental property of materials is the **specific heat** (c), which is the amount of heat that has to flow into an object for a unit temperature change per unit mass of the object. The SI unit is joules per kilogram per kelvin, or J/kg·K. In other units, it's calories per gram per degrees Celsius, or cal/g·°C.

- The equation for heat capacity is mass times specific heat, or $C = mc$, and then that product together is the heat capacity times ΔT, which equals the heat transferred to an object: $\Delta Q = C\Delta T$.

- Liquid water has very large specific heat, 4184 J/kg, to raise the temperature by just 1°C, which is why it's difficult to change the temperature of water.

- Suppose there's a power failure, and your house temperature drops to 5°C, which is about 41°F. How long will it take to warm your house to room temperature, 20°C? Your house's heating system supplies energy at a rate of 30 kW and has a heat capacity of 25 MJ/°C.

- The ΔT is up from 5°C to 20°C, so $\Delta Q = C\Delta T = 25$ MJ/°C × 15°C = 375 MJ of energy. The heat output rate is 30,000 watts— 1 watt equals 1 J/s—which is 30,000 J/s.

- The time will be the total amount of energy needed (375×10^6 J) divided by the rate at which energy is coming in (30,000 J/s), which equals 12,500 seconds, or about 3.5 hours.

- In terms of specific heat, heat flows are given by the masses of the objects times their specific heats times the temperature difference: $m_1 c_1 \Delta T_1 + m_2 c_2 \Delta T_2$, where the ΔTs are the changes in temperature of the 2 objects as they come to some final temperature. This equation characterizes the final equilibrium temperature.

- A blacksmith is going to plunge a 1.2-kg iron horseshoe at 650°C, which is about 1200°F, into 6 gallons of water, which is about 23 kg, at room temperature, 20°C. Neglecting any heat loss to the surroundings, what are the final temperatures of the water and the horseshoe?

- We can conclude that energy is conserved if no heat is lost, so the mass of the horseshoe times the specific heat of the horseshoe times the temperature change of the horseshoe plus the same values for the water equals zero: $m_h c_h \Delta T_h + m_w c_w \Delta T_w = 0$.

- Remember that one of the temperatures will increase and one will decrease, so one ΔT is positive and one is negative. Solving for T, we determine that $T = \dfrac{m_h c_h T_h + m_w c_w T_w}{m_h c_h + m_w c_w}$ and, by plugging in the numbers, that the water increases to 27°C.

Important Terms

absolute zero: The absolute limit of cold, at which all heat energy is removed from a system; equal to about −273°C.

heat: The kinetic energy (energy of motion) of the atoms or molecules making up a substance.

heat capacity: The amount of heat energy necessary to increase the temperature of a material by 1°C.

macroscopic properties: A generic term for phenomena and objects at the large scale. Everything that we can directly perceive may be regarded as macroscopic.

specific heat: The amount of heat that has to flow into an object for a unit temperature change per unit mass of the object. The SI unit is joules per kilogram per kelvin, or J/kg·K.

triple point: The point that defines a unique temperature and pressure for a substance at which its phases can coexist. For water, it is the point at which liquid water and solid water, or ice and water vapor, can coexist.

Suggested Reading

Rex and Wolfson, *ECP*, chap 12.1, 13.2.

Wolfson, *EUP*, chap 16.1–16.2.

Questions to Consider

1. Distinguish temperature and heat.

2. Otherwise identical, 2 objects are at different temperatures. Why isn't it appropriate to say that the hotter one contains more heat?

3. Does it take more, less, or the same amount of heat to raise the temperature of 1 kilogram of ice by 10°C as compared with 1 kilogram of liquid water?

4. The global temperature is expected to rise by about 3°C over the rest of this century. What's that in kelvins and in °F?

Heat Transfer

Lecture 22

T here are 3 important mechanisms of heat transfer: conduction, convection, and radiation. In all of these mechanisms, the rate of heat flow between an object and its surroundings increases with the temperature difference between object and surroundings. Supply energy to an object at a fixed rate, and its temperature will rise until the heat flow to its surroundings is equal to the rate at which energy is supplied. The object is then in thermal energy balance, and its temperature remains constant.

- Heat transfer is widely used in physics to describe the processes whereby heat flows. There are 3 important mechanisms of heat transfer—conduction, convection, and radiation—and these mechanisms have applications in technology and in nature.

- **Conduction** is a process of heat transfer between objects that are in direct contact. If the internal energy, or temperature, of an object is high, the molecules within the object are moving faster.

- For example, the rapidly moving molecules in a hot plate collide with the more slowly moving molecules in the walls of a beaker containing water that is placed on the hot plate. When a molecular transfer like this occurs, it tends to move energy preferentially from the faster-moving molecules (hotter) to the slower-moving ones.

- Eventually, the molecules in the beaker then transfer their energy to the inside wall of the beaker, where the energy is then transferred by conduction to the water.

- Heat transfer by conduction is a molecular process that requires 2 substances to be in direct contact; otherwise, heat can't flow through them.

- Conductive heat flow (H) is proportional to area (A), temperature difference (ΔT), and thermal conductivity (k), but it is inversely proportional to thickness (Δx): $H = -kA \dfrac{\Delta T}{\Delta x}$.

- Conductivity, therefore, depends on the area in question, the thermal conductivity, the temperature gradient or difference, and the thickness, or the distance through which the flow has to travel.

- Not surprisingly, the metals aluminum and copper are, along with being good conductors of electricity, good conductors of heat.

- One way the process of heat conduction is expressed, particularly in building materials, is with something called R-value, which measures the resistance to the flow of heat or material.

- The units of the R-value in the English system are heat squared times degrees Fahrenheit times hours per Btu—which stands for British thermal unit, the energy required to raise a pound of water by 1°F, or 1054 J.

- The inverse of the R-value is more meaningful and is called U, which is a kind of thermal conductive. It is 1 divided by the R-value and is a Btu per hour per square foot per degree Fahrenheit.

- **Convection** is the motion of energy, the flow of heat, by the motion of a fluid.

If a beaker is in contact with a hot plate, the electrical energy that heats the hot plate moves directly between the metal of the hot plate and the glass of the beaker and then into the water.

© Hemera Technologies/PhotoObjects.net/© Getty Images/Thinkstock.

- For example, once the heat conducted from a hot plate into a beaker reaches the bottom of the glass, the water is warm and expands, becoming less dense. The warm water then rises and carries its internal energy with it; the rising of that internal energy constitutes a temperature-driven flow of energy.

- When the water reaches the top, it is in contact with cooler air. It gives up its energy by conduction to the air, and then the water sinks down, forming patterns of convective motion that carry the heat around.

- Earth's atmosphere consists of many complicated cells of fluid motion that are caused by heating at the equator, where the Sun's rays hit most directly. Those cells of convective motion transfer energy toward the higher latitudes. If it weren't for those convective motions, the temperature difference between the equator and the poles would be much greater than it is.

- **Radiation** arises ultimately from accelerated electric charges. When there are thermal motions—that is, high internal energies or temperatures—there is also a lot of thermal agitation, which causes electromagnetic radiation.

- Radiation—light, infrared, ultraviolet—is the only mechanism of heat transfer that works in a vacuum.

- The amount of energy that radiates from a lightbulb, or at what rate it radiates energy, is given by something called the Stefan-Boltzmann law: $P = eA\sigma T^4$.

- P is the power emitted by a hot object through radiation; e is a quantity called the emissivity of the surface of an object; A is the surface area of the object that's radiating energy; σ is a universal constant, 5.67×10^{-8} W/m²·K⁴; and T is the temperature in kelvins.

- The temperature, which must be the absolute temperature, is raised to the 4th power, so the brightness of a lightbulb increases very rapidly as you increase the temperature. If you double the absolute temperature, the amount of energy increases by a factor of 2^4 or 16.

- The emissivity of the surface of an object is a number that ranges from zero to 1. If an object is completely black, it has a surface emissivity of 1, whereas if an object is completely shiny, its surface emissivity is zero.

- Earth is warmed by incoming sunlight; on average, sunlight reaches Earth's surface at the rate of about 240 W for every square meter of Earth's surface.

- The only way Earth can lose energy to space—because space is a vacuum—is through radiation. At the Earth's temperature, which is about 300 K, most of the radiation is invisible infrared.

- The outgoing radiation is established by the Stefan-Boltzmann law, which depends on Earth's temperature. The warmer Earth is, the rate of outgoing infrared is greater, so for Earth to have the condition of energy balance, the outgoing infrared has to equal the incoming sunlight.

- Earth is an object with emissivity approximately equal to 1, but you still need the total power coming to Earth from the Sun. You have the average power per square meter of surface, so you can divide $P = A\sigma T^4$ by A to get the power per area: $P/A = \sigma T^4$.

- In energy balance, power per unit area emitted in infrared radiation has to balance the power per unit area coming in as sunlight. Therefore, $P/A = 240$ W/m^2 if the planet is to be in energy balance.

- Solving for the temperature,

$$T = \sqrt[4]{\frac{240 \text{ W/m}^2}{\sigma}} = \sqrt[4]{\frac{240 \text{ W/m}^2}{5.67 \times 10^{-8} \text{ W/m}^2 \cdot \text{K}^4}} = \sqrt[4]{42 \times 10^8 \text{ K}^4}.$$

Then, $T = 255\ K = -18°C = 0°F$.

- The average temperature of the planet is 0°F. If it would've been 1000 K, Earth's oceans would have boiled away. If it had turned out to be $-100°C$, Earth's oceans would have frozen solid.

- An average of temperature of 0°F seems pretty cold, though. Earth's temperature is also affected by the greenhouse effect and the absorption of outgoing infrared by carbon dioxide in the atmosphere. That is a complicated story that explains why Earth's climate is, in fact, changing.

Important Terms

conduction: Heat transfer by physical contact.

convection: Heat transfer resulting from fluid motion.

radiation: Heat transfer by electromagnetic waves.

Suggested Reading

Rex and Wolfson, *ECP*, chap 13.4.

Wolfson, *EUP*, chap 16.3–16.4.

Questions to Consider

1. Conduction, convection, and radiation all play roles in establishing Earth's climate. But our planet's loss of energy to outer space is ultimately by radiation only. Why?

2. You go to a home supply store and purchase some R-19 insulation. Describe quantitatively what R-19 means.

3. When sunlight shines on a solar greenhouse, why doesn't the greenhouse temperature continue to rise as long as the Sun is shining on it?

Matter and Heat
Lecture 23

H eat flow into a substance usually raises its temperature, but it can have other effects, including so-called thermal expansion and changes among solid, liquid, and gaseous forms—collectively called phase changes. Substantial energy is required to change phase, and temperature remains constant while the phase change is occurring. Diagrams of pressure versus temperature—phase diagrams—describe the details. Water's thermal behavior is unusual and has important consequences for aquatic life.

- Matter responds to heat flows in 2 important ways: It can change shape or size—expand or contract as heat flows in or out—or change phase—from a solid to a liquid or from a liquid to a gas, and vice versa.

- The internal energy that increases when heat flows into a material is actually the kinetic energy of its molecules and their thermal agitation as they're bouncing around. As the molecules warm up and bounce faster, they effectively increase the mean intermolecular distance, causing **thermal expansion**.

- Thermal expansion is characterized by the **coefficient of thermal expansion**, which is the fractional change that an object undergoes as a result of a temperature change of 1°.

- We can define the coefficient of length expansion in the case of solids as $\alpha = \frac{\Delta L}{L} / \Delta T$ and the coefficient of volume expansion in the case of liquids and gases as $\beta = \frac{\Delta V}{V} / \Delta T$.

- One aspect of thermal expansion that is anomalous is the thermal expansion of water. Water is a very important but unusual substance; it is one of the few substances whose solid state is actually less dense than its liquid state.

- Water floats because the molecular structure of water ice is open, and the crystalline structure allows a lot of empty space, which causes the density to decrease. From 0°C to 4°C, liquid water still contains a residue of the bonding that formed its molecular structure, and as that boding breaks (as the water is warmed), the density increases with increasing temperature.

- One thing that happens to matter when heat flows into it is it expands or, in the case of water, it contracts in the realm of 0°C to 4°C.

- When heat flows into matter, it can also undergo phase changes, during which 2 phases actually coexist. For example, when ice is melting, both ice and water are present. During that melting, the temperature does not change.
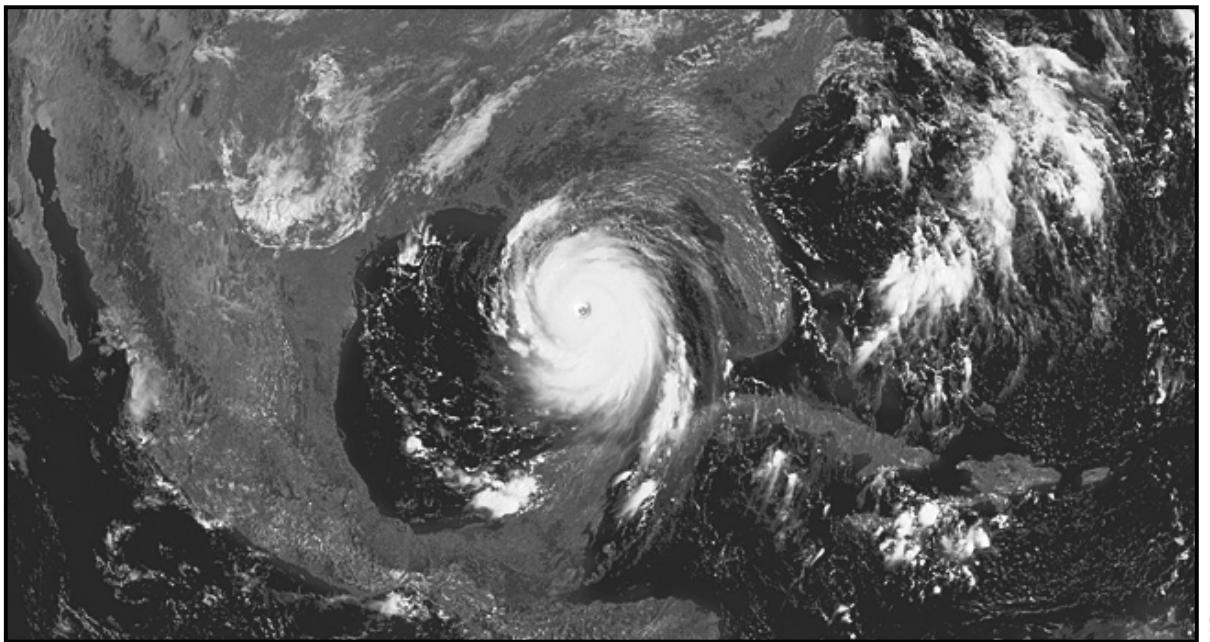
NOAA.

Hurricanes are driven by the energy associated with water that has been evaporated from the warm, tropical ocean into the atmosphere.

- Phase changes require energy because the process includes the breaking of bonds that hold molecules together.

- In the case of ice, its bonds lock the water molecules into a rigid crystalline structure. In order to turn that material into a liquid in which the molecules are still fairly tightly bound together but are able to move around each other, those bonds must be broken, and that requires energy.

- It takes additional energy to break those weaker bonds that are holding the material into a dense liquid and allow the molecules to go off on their own as they do in a gas.

- The energies that it takes to transform from one phase to another are called **heats of transformation**—the heat of fusion and the heat of vaporization are examples.

- The total energy (Q) required to change the phase of a mass m is $Q = Lm$, in which L is the energy per unit mass needed to cause phase changes.

- Water has a very large heat of vaporization. The boiling point of water is 100°C, or 373 K, and water's melting point is 0°C, or 273 K.

- The process of heat of transformation associated with vaporized water is one of the major mechanisms for obtaining energy from Earth's surface, where the Sun deposits most of its energy, into the atmosphere.

- A **phase diagram**, like the one in Figure 23.1, plots the various phases of a substance, with temperature on the horizontal axis and pressure on the vertical axis. For a typical substance, the curves of the graph divide the space into 3 regions: solid, liquid, and gas.

Figure 23.1
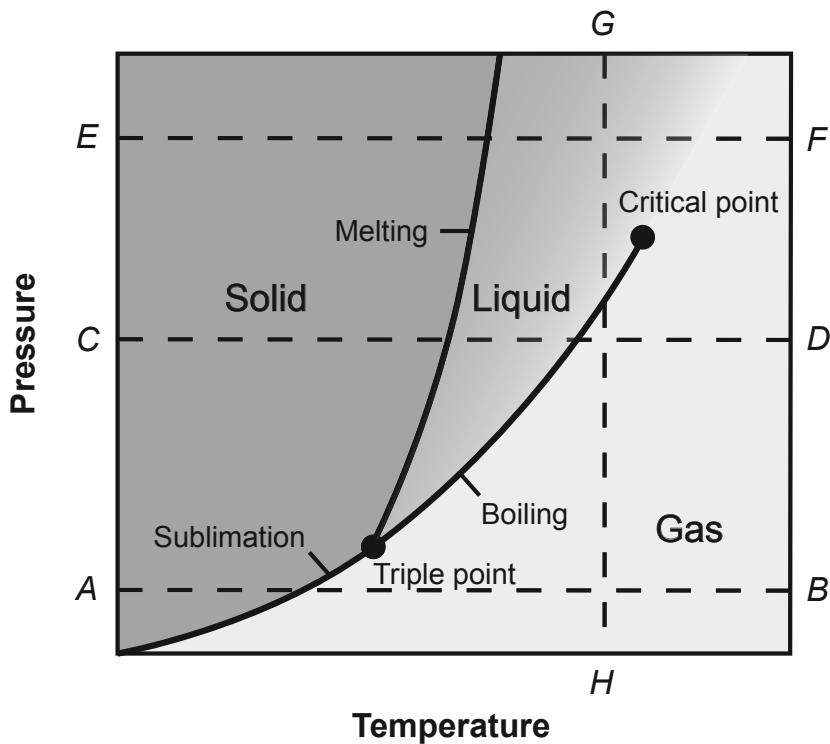
- The triple point, as defined in Lecture 21, is the unique point at which all 3 phases can simultaneously coexist. This point has a unique temperature and pressure at which the 3 can coexist.

- For water, the triple point occurs at a temperature of 273.16 K and a pressure of about 6 thousandths of an atmosphere. The temperature scale is defined by the triple point of water, 273.16, and absolute zero, 0.

- There are 3 processes that change the phases of materials: A few common processes are melting—changing from solid to liquid— and boiling—changing from liquid to gas. A less common process is sublimation, which is the process of changing a solid directly into a gas.

- The process of sublimation does not occur for water under ordinary atmospheric conditions, but it is common for carbon dioxide, for example. Below the triple point of carbon dioxide, it does not exhibit a liquid phase—it changes directly into a gas.

- In Figure 23.1, a substance could move from point E to F at relatively high pressure and cross the solid-liquid transition, but it would never encounter the solid-gas transition. Above the critical point, there's just a dense fluid phase that gradually transforms from being more liquid-like to more gaseous-like with the abruptness of a phase transition that occurs at a particular point called the **superfluid** state.

- Because water is a fascinating substance that behaves anomalously relative to most everyday substances, the phase diagram for water is a bit different.

- The solid-liquid curve goes up to the left instead of going up to the right as in Figure 23.1. This is a reflection of the fact that the density of water actually increases for a short time after it turns from solid to liquid.

- This property of water allows a process that cannot be accomplished with other substances, and it's called **pressure melting**: If you increase the pressure on water when it's in the solid phase, it will turn into a liquid as it crosses the solid-liquid boundary.

- For most substances, if you increase the pressure on a liquid at a fixed temperature, the substance will cross the boundary into a solid. However, water crosses the boundary from solid into liquid.

Important Terms

coefficient of thermal expansion: The fractional change that an object undergoes as a result of a temperature change of 1 degree.

heat of transformation: The energy that it takes to transform from one phase to another; for example, the heat of fusion and the heat of vaporization.

phase diagram: A diagram showing how the phases of a substance relate to its temperature and pressure.

pressure melting: A unique property of water that occurs when the pressure on water in the solid phase is increased, causing it to turn into a liquid as it crosses the solid-liquid boundary.

superfluid: A liquid at extremely low temperatures that has many surprising properties, including zero viscosity.

thermal expansion: As internal energy increases when heat flows into a material, the mean intermolecular distance increases, resulting in pressure or volume changes.

Suggested Reading

Rex and Wolfson, *ECP*, chap 12.2, 13.3.

Wolfson, *EUP*, chap 17.2–17.3.

Questions to Consider

1. Water's thermal expansion properties are unusual. Why, and in what temperature range? What are some consequences?

2. Why isn't the boiling point of water a suitable standard for calibrating thermometers, whereas the triple point is?

3. Can boiling water exceed 100°C (212°F)? Explain.

The Ideal Gas
Lecture 24

A particularly simple system for understanding thermal behavior is the ideal gas, comprising widely spaced, noninteracting molecules. Theoretical analysis of the ideal gas using Newtonian mechanics reveals that temperature is a measure of molecular kinetic energy. Consequently, temperature is also a measure of the typical molecular speed. Real gases under ordinary conditions obey the ideal-gas law with remarkable accuracy, and the ideal-gas law encompasses not only idealized point-like molecules, but also real, complex molecules that can rotate and vibrate.

- The behaviors of gasses provide deep insights into some fundamental laws of thermodynamics that we'll be dealing with in upcoming lectures.

- The **ideal gas** is simple and easy to characterize primarily because its molecules are far apart and exhibit very few interactions—they only occasionally undergo collisions.

- Other materials, such as solids and liquids, are much more difficult and complicated to understand—as are gasses that exhibit nonideal behavior.

- In the realm where the ideal gas approximation applies, gasses behave basically universally, regardless of the nature of their molecules and regardless of the type of gas.

- Experiments with gasses show that if you fix the temperature of a gas, then the pressure of the gas is inversely proportional to its volume. The product of the pressure and the volume, pV, remains constant at a fixed temperature.

- If you change the temperature, the quantity pV changes proportionately. If you change the amount of gas, pV also changes proportionately.

- All of these gaseous properties culminate in the **ideal-gas law**, which says that the pressure of a gas times the volume that gas occupies is the product of the number of molecules (the amount of gas), a constant of nature, and the temperature: $pV = NkT$.

- Pressure is p; volume is V; N is the number of molecules; k is called **Boltzmann's constant**, which is a conversion between temperature and energy (1.3×10^{-23} J/K in SI units); and T is the temperature, which has to be in absolute units and in kelvins.

- There are a few assumptions we must make when dealing with ideal gasses, starting with the fact that they are in a closed container, whose volume determines the volume of the gas.

- We also assume that the gasses are composed of identical molecules, and each molecule has mass m. There is no internal structure or size to these molecules.

- A critical assumption is that the molecules do not interact by long-range forces—they do not attract or repel each other at a distance.

- If these assumptions are violated, then the gas does not behave ideally. To some extent, all gasses violate these assumptions slightly, but the violations are so small that they often do not affect the results.

- Let's also assume that there aren't any collisions between molecules to simplify the math—this is not an essential assumption.

- Unless we're dealing with a strange situation that sets up some preferred direction in a system, we also assume that the molecules have a random distribution of velocities.

- Gravity is so weak and molecules are moving so fast at typical temperatures that gravity is not an issue.

- Using Newtonian mechanics, we can assume that the collisions with the container walls are elastic, which conserves kinetic energy. Therefore, when a molecule barrels into the wall of the container, it doesn't transfer any of its energy to the wall.

- The end of the container, as shown in Figure 24.1, has a surface area A, and the container has a length L. The molecules are traveling in many directions, but we will label a direction x, which will be used to generalize the other ways the molecules are traveling.

Figure 24.1

Surface area A

L x
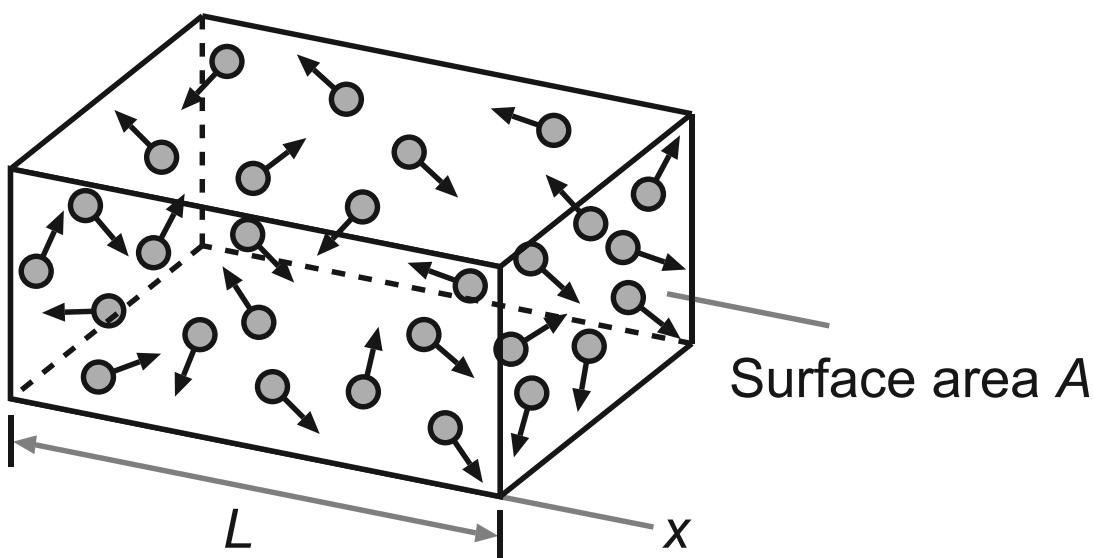
- One molecule, which we will designate as the ith molecule, has mass m and an x-component of its velocity, which is v_{xi} because it's the x-component of the i molecule's velocity.

- The ith molecule collides with the end of the container in the x direction. When it does, it exerts a force. It hits the end wall at the maximum x value in this container, and it bounces off at the same angle it came in with to conserve momentum.

- Using Newtonian mechanics, because the molecule bounces back, its x-component of motion reverses—changes direction—which means it has not only given momentum to the wall, it has also gained momentum from the wall in that collision. The momentum change is twice its initial momentum: $\Delta\mathbf{p}_i = 2mv_{xi}$.

- The time until the molecule returns to the wall after its collision with the wall and traveling to the other end of the container is twice the distance divided by speed: $\Delta t = 2L/v_{xi}$.

- The average force on the surface area A the ith molecule exerts is the change in momentum with respect to time:
$$\mathbf{F}_i = \frac{\Delta\mathbf{p}_i}{\Delta t} = \frac{2mv_{xi}}{\left(2L/v_{xi}\right)} = \frac{mv_{xi}^2}{L}.$$

- The average force on the wall due to all N molecules is the sum, as denoted by the Greek letter sigma (Σ), of these forces over all the values of i: $\mathbf{F} = \sum_i \frac{mv_{xi}^2}{L} = \frac{m}{L}\sum_i v_{xi}^2$.

- We factored out the m because it's the same for all molecules, and certainly the length of the container is the same.

- Pressure is defined as the force per unit area, and we'll use the average force: $p = \frac{\mathbf{F}}{A} = \frac{m}{AL}\sum_i v_{xi}^2 = \frac{m}{V}\sum_i v_{xi}^2$.

- Multiplying the right-hand side by 1 in the form of N/N, we get $p = \frac{Nm}{V}\sum_i \frac{v_{xi}^2}{N} = \frac{Nm}{V}\overline{v_x^2}$, where $\overline{v_x^2}$ is the average of the squares of the x-velocity components.

- Molecular motion is random, so the average of the x velocity squared and the other velocities squared has to be the same: $\overline{v_x^2} = \overline{v_y^2} = \overline{v_z^2}$.

Lecture 24: The Ideal Gas

- Because the x and y and z directions are all at right angles, then these velocities add by the Pythagorean theorem: $\overline{v^2} = \overline{v_x^2} + \overline{v_y^2} + \overline{v_z^2}$.

- Therefore, $\overline{v^2} = 3\overline{v_x^2}$, and $p = \dfrac{Nm}{3V}\overline{v^2}$.

- Multiplying both sides by the volume, V, we get: $pV = \dfrac{1}{3}Nm\overline{v^2}$.

- Multiplying the right-hand side by 1 in the form 2/2, we get: $pV = \frac{2}{3}N\left(\frac{1}{2}m\overline{v^2}\right)$.

- We are now left with $pV = 2/3N$ times a term, which is the average kinetic energy. Comparing this equation with the ideal-gas law, both have pV on the left-hand side and NkT in different forms on the right-hand side.

- The huge insight we get from this is that temperature is a measure of the average molecular kinetic energy, which is the theory of the ideal gas: $\frac{1}{2}m\overline{v^2} = \frac{3}{2}kT$.

- Complex molecules still behave as ideal gasses, but their specific heats, the amount of energy it takes to raise the gas temperature, vary and depend on the structure of the molecule. The molecule can have a number of ways that energy can be divided in different modes of motion or potential energy in those molecules.

- The ideal-gas law is a great approximation that applies to real gasses in the real world to a very good extent, but as you begin to approach the point at which gasses liquefy, that approximation breaks down—first in minor ways, and then in more major ways.

Important Terms

Boltzmann's constant: A conversion between temperature and energy. In SI units, it is 1.3×10^{-23} J/K.

ideal gas: A theoretical gas that contains molecules that are far apart and exhibit very few interactions. In the realm where the ideal gas approximation applies, gasses behave basically universally, regardless of the nature of their molecules and regardless of the type of gas.

ideal-gas law: The pressure of a gas times the volume that gas occupies is the product of the number of molecules (the amount of gas), a constant of nature, and the temperature: $pV = NkT$.

Suggested Reading

Rex and Wolfson, *ECP*, chap 12.3–12.4.

Wolfson, *EUP*, chap 17.1.

Questions to Consider

1. To what more fundamental quantity does the temperature of an ideal gas correspond?

2. Why does the ideal-gas approximation break down at high gas densities?

Heat and Work
Lecture 25

T he first law of thermodynamics is a statement of energy conservation in the context of thermodynamics. The first law states that a system's internal energy can be changed either by heat flow to the system, by doing mechanical work on the system, or by a combination of the 2. Although there are infinitely many ways to combine work and heat, specific thermodynamic processes—including isothermal and adiabatic processes— prove useful for understanding thermal energy flows. These processes are particularly simple to analyze when applied to ideal gases.

- This lecture discusses things that can be done to ideal gasses that changes their internal energy or effects other changes in them.

- Lecture 11 introduced the conservation of mechanical energy and found that when only conservative forces act—forces that give back the energy stored as potential energy—then the sum of the kinetic and potential energy doesn't change.

- Earlier examples explicitly excluded processes like friction that involve nonconservative forces because mechanical energy is lost—but, actually, it just changes into internal energy.

- Therefore, we can generalize the principle of conservation of energy to include internal energy—also called thermal energy—and give it the symbol *U*, which is not to be confused with the *U* used for potential energy in mechanics.

- The **first law of thermodynamics** includes the effects of nonconservative forces that generate internal energy and says that there are 2 ways to change the internal energy of a system: by adding heat to the system or by doing work on the system.

- The first law of thermodynamics states that the change in a system's internal energy, ΔU, is equal to any heat (energy in transit because of a temperature difference) that may have flowed into the system: $\Delta U = Q + W$ (both Q and W are positive because heat and work are being added to the system).

- We're going to use the ideal-gas system because its internal energy is a direct measure of temperature due to the fact that there's no interaction between the molecules.

- A particular ideal-gas system consists of a cylinder and a piston that is free to move up and down in that cylinder. You can make changes to the volume, temperature, and pressure of the existing gas or even add additional gas to it.

- In this case, the cylinder is thermally insulated on all sides, and the piston is also insulated, so heat can't flow in and out. The bottom of the cylinder may or may not be insulated so that we can allow or prevent heat flow.

- Figure 25.1 is a *pV* **diagram**—volume is on the horizontal axis and pressure is on the vertical axis—which shows the relationship among pressure, volume, and temperature for an ideal gas.

Figure 25.1

p_1, V_1, T_1

p_2, V_2, T_2
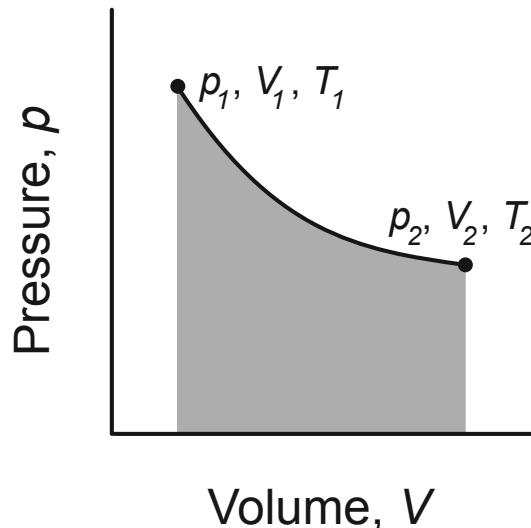
Pressure, p

Volume, V

- Starting at pressure p_1 and volume V_1, there will be some temperature T_1, and given the number N of molecules, the ideal-gas law ($pV = NkT$) would tell us what the temperature is. Any 2 of those quantities fix the third.

- By gradually changing pressure and volume, for example, and moving the gas sample to a different point in its pV diagram, the entire gas will stay in equilibrium and will have a well-defined temperature. The gas goes through a sequence of states in its pV diagram, and if this occurs slowly enough, it becomes a reversible process in which the gas could then go back through the same sequence of states.

- Piston-cylinder systems have practical applications. For example, gasoline engines are the kinds of devices in which pistons are moving up and down in cylinders, and work is being done. These systems are also involved with meteorology.

- The piston has area A, and there is a pressure of the gas p, which is force per area, so the gas exerts a force p times A on the piston ($F = pA$). The piston rises a distance Δx, and work is force times distance, so the work done by the gas is $W = F\Delta x$.

- The force the gas is exerting times the distance it's moved is $pA\Delta x$, in which $A\Delta x$ is just the volume change. The gas's volume has also changed by an amount ΔV.

- The work done by the gas (W), assuming that it didn't move very far so that the pressure stayed constant, is $p\Delta V$. If the pressure didn't stay constant, then the work done by the gas would be the area under the pV curve.

- There are many thermodynamic processes, and one of them is called an **isothermal process**—isothermal means that the temperature is the same—which occurs at a constant temperature.

- To make an isothermal process occur in an ideal gas, it needs to be in contact with a heat reservoir, maintained at a fixed temperature T, so the temperature of the gas will always be the temperature of the reservoir. For example, a hot plate and a huge volume of water are heat reservoirs.

- The first law of thermodynamics says that the change in internal energy is $Q + W$. In an ideal gas, the internal energy is proportional to the temperature, so ΔU is zero for an isothermal process in an ideal gas, so $Q + W = 0$, or $Q = -W$.

- If heat flows into the system, Q is positive, so W will be negative, and the system will do work on its surroundings. The gas will expand as heat flows into it, and it will push up on the piston. In other words, negative work will have been done on the gas. The opposite is also true when Q is negative.

- In an isothermal process, the work done can be found by analyzing the ideal-gas law, $pV = NkT$, at a constant T so that p becomes proportional to $1/V$, which causes a hyperbolic path in the pV diagram.

- The work done on the gas depends on the amount of gas, the temperature T, and the logarithm of the ratio of the initial to final volume.

- If you work out whether that logarithm is positive or negative, depending on whether V_1 is bigger or smaller than V_2, it gives you the work done on the gas: If that number comes out positive, work is actually done *on* the gas. If it comes out negative, work is done *by* the gas.

- In an **adiabatic process**, there's no heat flow, which means that no heat is allowed to flow into or out of the system. You can achieve this by insulating the system from its surroundings or by making the process go so rapidly that there's no time for heat to flow.

- Calculus shows that pV raised to some exponent called the adiabatic exponent, gamma (γ), which is characteristic of a given gas, is a constant in adiabatic processes.

- The first law of thermodynamics for an adiabatic process states that the change in internal energy is $Q + W$, but there's no change in internal energy in an adiabatic process, so Q is zero, which tells us that $\Delta U = W$.

- If you do work on a system, the internal energy—and therefore the temperature if it's an ideal gas—increases for adiabatic processes.

Important Terms

adiabatic process: A process that takes place without any exchange of heat with its surroundings, during which entropy remains constant. If a gas undergoes adiabatic expansion, its temperature decreases.

first law of thermodynamics: The statement that energy is conserved, expanded to include thermal energy.

isothermal process: A process that takes place at a constant temperature. A gas that undergoes isothermal expansion will have to absorb heat from its surroundings.

***pV* diagram**: A diagram in which volume is on the horizontal axis and pressure is on the vertical axis that shows the relationship among pressure, volume, and temperature for an ideal gas.

Suggested Reading

Rex and Wolfson, *ECP*, chap 14.1–14.2.

Wolfson, *EUP*, chap 18.

Questions to Consider

1. You could raise the temperature of a jar of water a given amount ΔT by heating it over a flame, in a microwave oven, or by shaking it violently. Would there be any difference in the thermodynamic properties of the water depending on which method you used?

2. In an isothermal process, pV is constant, but in an adiabatic process, $(pV)^{\gamma}$ is constant. Why aren't these 2 equations contradictory?

3. Does the trapping of pollutants under adverse atmospheric conditions require that temperature actually rise with height? If not, what is the criterion?

Entropy—The Second Law of Thermodynamics
Lecture 26

I n its broadest form, the second law of thermodynamics asserts the universal tendency of systems toward disorder. Entropy is a measure of disorder, and the second law states that entropy generally increases—and, in any case, can't decrease. Creating order out of disorder requires the expenditure of energy to do the work of creating a more ordered state, and obtaining that energy means increasing disorder elsewhere—so entropy still increases.

- Unlike the other laws of physics addressed so far, the **second law of thermodynamics** talks about what is likely to happen or, more importantly, what is unlikely to happen—it has a probabilistic side to it.

- The second law of thermodynamics is crucially important in understanding the workings of the world—not only in physics, but in biology, chemistry, geology, and evolution.

- The second law of thermodynamics is about processes that are so rare, so improbable, that in our real world they simply never happen.

- For example, the process of beating an egg is completely irreversible. In principle, however, it could be reversed—there is nothing that violates any fundamental law of physics—and yet it just doesn't occur.

- Organized states are rare compared to disorganized states. Once the system gets into a disorganized state, there are many other disorganized states it could transit into, but there are very few organized states, so it's unlikely to transit into a more organized state. This tendency toward chaos and disorder is at the heart of the second law of thermodynamics.

- **Entropy** is a measure of the disorder in a system. Entropy tends to increase over time, and an entropy increase in a system is accompanied by a decrease in the ability of that system to do useful work.

The concept of entropy is that systems tend naturally toward disordered states.

- For example, there is a closed box that's insulated from its surroundings, so no heat can flow in or out. In the box, there is a partition. On one side of the box, there's a gas that has some volume, pressure, and temperature, and the volume of the gas is exactly half the volume of the box. On the other side of the box is a vacuum.

- When you remove the partition, what happens? The gas spreads itself evenly throughout the entire box, and eventually there is just a random hodgepodge of molecules rushing around.

- Because the box has not expanded and its volume hasn't changed, the gas has done no work—it still has the same amount of internal energy, and the temperature is still the same. This system is in a state of adiabatic free expansion that is also isothermal.

- A **microstate** is a specific arrangement of individual molecules. A **macrostate** is characterized by the number of molecules that are on each side of the box.

- For a 2-molecule gas, there are 4 possible microstates (2^n): They could both be on the left side of the box, both on the right side of the box, one on the right and one on the left, and then the 2 could switch sides.

- For this same 2-molecule gas, there are only 3 macrostates ($n + 1$): There are the states where all the molecules are together on one side, where they're together on the other side, and where they're split evenly. There are more microstates than there are macrostates.

- If this gas gets itself randomly into some state, we're more likely to find it in the macrostate where there are equal numbers on both sides because there are 2 corresponding microstates that are equivalent to that macrostate, which allows for an increased probability for the molecules to be in that position.

- A typical gas might have 10^{23} molecules, so it has $2^{10^{23}}$ microstates and $10^{23} + 1$ macrostates. Essentially, half the molecules would be on one side of the hypothetical box or room, and half the molecules would be on the other side. It is just so improbable that it's never going to happen that they're all on one side.

- The change in entropy is defined for a reversible process—a process that has well-defined pressure, volume, and temperature—as the heat flow divided by the temperature at which this process occurs:
$$\Delta S = \frac{Q}{T_2} - \frac{Q}{T_1}.$$

- In terms of entropy, the second law of thermodynamics says the entropy of a closed system can never decrease—at best, it can remain constant.

- In a closed, insulated container, hot water, T_h, and cold water, T_c, come to some equilibrium temperature—some intermediate temperature that is in between the hot temperature and the cold temperature, a warm temperature T.

- The average temperature of the hot water lies somewhere between the initial temperature of the hot water (T_h) and the final warm temperature they both share (T). Similarly for the cold water, there's an intermediate temperature T_2 that lies somewhere between T_c and T.

- The entropy change of the hot water is $-Q/T_1$ because in the definition of entropy, $\Delta S = Q/T$, Q is the heat that flowed into a system. There was a heat flow to the cold water, so energy flowed out of the warm water as a flow of heat, and the heat Q is negative in this case.

- The entropy change of the cold water is $+Q/T_2$, which is the temperature that characterizes the average for the cold water. Because heat flowed in to the cold water from the hot water, the Q is positive.

- The Q is the same in both cases of cold and hot water because the entire system is in an insulated container, so there is no energy loss.

- The entropy of the hot water decreased, and the entropy of the cold water increased. There's nothing about the second law of thermodynamics that says entropy can't decrease; instead, it says the entropy of a closed system can't decrease. Therefore, the increase will be greater than the decrease.

- The change in the system's entropy is $\Delta S = \dfrac{Q}{T_2} - \dfrac{Q}{T_1} > 0$.

- This quantity is greater than zero because T_2 is lower than T_1, so the subtraction yields a positive number. Quantitatively, this is a description in a simple system of how entropy increases.

- Life and civilization are both examples of systems that grow more organized, but they can do so only with an external energy source: the Sun. The generation and outpouring of energy from the Sun increase disorder more than life and civilization decrease it, so the second law is satisfied.

- Whenever you encounter a process that appears to decrease in entropy, you're not dealing with a closed system because any entropy-decreasing process is offset by an entropy-increasing process.

Important Terms

entropy: A quantitative measure of disorder. The second law of thermodynamics states that the entropy of a closed system can never decrease.

macrostate: A state characterized by the number of molecules that are located on each side of a divided area.

microstate: A specific arrangement of individual molecules.

second law of thermodynamics: A general principle stating that systems tend to evolve from more-ordered to less-ordered states.

Suggested Reading

Rex and Wolfson, *ECP*, chap 14.3, 14.5.

Wolfson, *EUP*, chap 19.1–19.2, 19.4.

Questions to Consider

1. What is entropy?

2. How does the second law of thermodynamics differ from a strictly deterministic law like Newton's second law?

3. In reference to the second question for Lecture 25, how can the adiabatic free expansion described in this lecture be both adiabatic and isothermal?

Consequences of the Second Law
Lecture 27

T he random molecular motions associated with thermal energy represent a disordered state of relatively high entropy. As a result, the second law of thermodynamics precludes converting random thermal energy into more organized energy. However, we can build heat engines that convert some random thermal energy into useful mechanical or electrical energy. The second law puts explicit limits on the efficiency of heat engines and on the ability of refrigerators to extract thermal energy to provide cooling.

- In the last lecture, we learned that you can't convert random thermal energy directly into mechanical energy with 100% efficiency.

- Nevertheless, we run a lot of our world on heat engines—devices that do extract energy from hot sources such as burning gasoline, burning oil, burning coal—that convert heat into mechanical energy.

- The second law of thermodynamics says these machines can't extract energy with 100% efficiency, but they do.

- Lord Kelvin and Max Planck interpret the second law of thermodynamics as saying that it is impossible to build a heat engine operating in a cycle that extracts heat from a reservoir, or a source of thermal energy, and delivers an equal amount of work to the amount of heat extracted.

- What they're saying is that, in a cycle, you can briefly extract energy from heat and turn it all into work, but overall, the system cannot turn heat into an equal amount of energy. Their statement is a more restricted version of the statement that entropy must increase.

- A heat engine extracts energy from a hot reservoir and delivers some mechanical work, but—according to the second law of thermodynamics—it has to reject some energy as a heat flow to a lower-temperature reservoir, typically to the surrounding environment.

- Cars have radiators to get rid of waste heat, and power plants have cooling towers to get rid of waste heat. A perfect heat engine, one that does not waste heat, is impossible.

- In the 19th century, French engineer Sadi Carnot worked out the details of a particular engine called a Carnot engine and its efficiency. Real engines aren't built exactly like Carnot engines—in principle, they could be—but their workings are somewhat similar.

- **Carnot engines** are cyclic and, in principle, reversible. Because its operations take place very slowly, systems stay in thermodynamic equilibrium, so the paths in *pV* diagrams could be reversed.

- A Carnot engine could take in work and transfer heat from a cool substance to a hot substance, which is exactly what a refrigerator does.

- The Carnot engine has a hot reservoir and a cold reservoir. It also has some mechanism for extracting work from the hot reservoir, namely a piston-cylinder system, which is sometimes connected by a crankshaft to a wheel that can be turned to do mechanical work.

- A crude heat engine needs a hot and a cold reservoir. Not all the energy extracted from the hot reservoir ends up as mechanical work, but some of it does. The rest is dumped as waste heat into the cool reservoir.

- In fact, if the cool reservoir weren't big enough, it would gradually heat up, and you wouldn't be able to run the engine anymore. The 2 waters would come to the same temperature, which leads to increasing entropy and the loss of ability to do work.

- In any kind of device, efficiency is a measure of what we would like that device to give us—in the case of an engine, we want energy (mechanical work) versus what we have to put in (the energy content of fuel).

- For a heat engine, we are putting in energy extracted from a hot source in the form of a heat flow (the hot reservoir, which in practice would be the burning of fuel). The efficiency is the amount of work we get out divided by the amount of heat we had to put in: $e = W/Q_h$.

- The work is the difference between the heat that is dumped into the cold reservoir and the heat flow that doesn't go to the cold reservoir, which ends up as work. If the engine operates in a cycle, its internal energy doesn't change over a whole cycle.

- Using the first law of thermodynamics ($\Delta U = Q + W$), there's no change in internal energy, so the net heat, $Q_h - Q_c$, is the work we get out. Therefore, instead of W, we can substitute $Q_h - Q_c$.

- Algebraically, the efficiency of the engine is $e = \dfrac{Q_h - Q_c}{Q_h} = 1 - \dfrac{Q_c}{Q_h}$.

- From the isothermal and adiabatic relationships introduced in Lecture 25, you find that the ratio of heat is the ratio of the cold temperature to the hot temperature: $\dfrac{Q_c}{Q_h} = \dfrac{T_c}{T_h}$.

- Therefore, the maximum possible efficiency of Carnot's engine, assuming there is no friction, is $e = 1 - \dfrac{T_c}{T_h}$.

- Another consequence of the second law of thermodynamics—another thermodynamic impossibility—is that heat doesn't flow spontaneously from cold to hot, which means that it's impossible to build a refrigerator because that's what a refrigerator does.

- If you put a lukewarm item in the refrigerator, it gets cold, but you probably don't notice that heat comes out of the back of the refrigerator that was extracted from the item.

- The statement is saying that you can't just extract the heat from the item and dump it out the back of the refrigerator—that this won't happen spontaneously.

- Instead, what has to happen is you also have to do mechanical work, which is why you have to plug your refrigerator in. A real refrigerator works just like an engine in reverse.

- The Clausius statement, an interpretation of the second law by Rudolph Clausius, says that it's impossible to build a perfect refrigerator—just like the Kelvin-Planck statement said it was impossible to build a perfect engine.

- **Carnot's theorem** states that it is thermodynamically impossible to build an engine whose efficiency is better than a Carnot engine.

- The efficiency of real engines is actually worse for several reasons. First, if any process occurs, it's irreversible. In other words, if the system ever leaves thermodynamic equilibrium, various nonequilibrium situations (like friction) reduce the efficiency.

- Another problem is if you don't maintain a fixed temperature of the hot and cold reservoirs, then the Carnot efficiency involves some kind of average temperature, so you may be getting hotter temperatures than you need.

- The Carnot efficiency is the absolute maximum limit for any real engine. In principle, we can get arbitrarily close to it, but in practice, we don't do very well.

- Energy has not just quantity, but also quality. Energies associated with low temperatures (as compared to their surroundings) are situations of high entropy and low energy quality.

- Higher temperatures relative to the surroundings have higher energy quality because a more efficient heat engine can run. The absolute highest quality of energy comes from mechanical energy and electrical energy, which can convert to any form with 100% efficiency; thermal energy, on the other hand, is a low-quality energy.

- Therefore, it is a waste to use high-quality forms of energy, like electricity and mechanical energy, for heating because you're turning the same amount of energy into energy of very low quality.

Important Terms

Carnot engine: A simple engine that extracts energy from a hot medium and produces useful work. Its efficiency, which is less than 100%, is the highest possible for any heat engine.

Carnot's theorem: A theorem named after French scientist Sadi Carnot that states it is thermodynamically impossible to build an engine whose efficiency is better than a Carnot engine.

Suggested Reading

Rex and Wolfson, *ECP*, chap 14.4.

Wolfson, *EUP*, chap 19.3.

Questions to Consider

1. Why can't engineers make power plants nearly 100% efficient? Or is it not their fault? Explain.

2. In what way are engines and refrigerators opposite devices?

3. Electric waters are inexpensive and nearly 100% efficient in converting electrical energy to the internal energy of hot water, yet they make less sense thermodynamically than gas water heaters. Explain.

A Charged World
Lecture 28

Electric charge is a fundamental property of matter that comes in discrete units; the elementary charge is equal to the proton's charge. The net charge in a closed region cannot change—even though positive and negative pairs can be created or annihilated. Like charges repel, while opposites attract. In either case, the electric force depends on the product of the charges and inversely on the square of the distance between them. The superposition principle allows us to find the electrical influence of distributions of electric charge.

- **Electromagnetism** is the force that dominates on scales from roughly the size of the atomic nucleus to objects the size of human beings. Beyond that, gravity begins to become more important than electromagnetism.

- Electrical forces bind atoms into molecules and molecules into biological tissues. Electromagnetic technology is the basis of everything from giant motors that run subway trains to processes of the nanoelectronic scale.

- The ancient Greeks learned about electricity through their study of the substance amber, which is a very good electrical insulator that easily builds up static electricity.

- The Chinese, at about the same time, experimented with electromagnetism and developed magnetic compasses.

- In the 18th century, Benjamin Franklin proposed a model of electricity in which he envisioned it as a kind of fluid. He was the first to realize that there were probably 2 kinds of this fluid, which we call 2 different charges today.

- Joseph Priestley in England and Charles Coulomb in France both quantified the electric force by discovering an equation that described how the electric force between 2 charges behaved.

- Luigi Galvani and Alessandro Volta developed the first battery in the 1800s. They studied electric currents, often generated in biological systems at first.

- By the time the 19th century came around, people were beginning to understand electricity and magnetism as related things. Hans Christian Oersted and André-Marie Ampère determined some of the relationships between electricity and magnetism, and Michael Faraday discovered the phenomenon of electromagnetic induction.

- James Clerk Maxwell completed electromagnetic theory in the 1860s; the classical physics version of electromagnetism became complete at that time.

- **Electric charge** is a fundamental property of matter. There are 2 kinds of electric charge, as Benjamin Franklin suggested—positive and negative. Like charges repel, and opposite charges attract.

- Electric charge is a conserved quantity. If you have a closed region, the total amount of electric charge in that region can never change.

- Electric charge is even more conserved than mass: Particles can appear and disappear, but net electric charge, net electric charge, can't change. A positive and negative charge can appear, but that's still zero net charge.

- Charge is quantized: It comes in small, discrete amounts. In 1909, Robert Millikan discovered the elementary charge, and it's given the symbol e.

- We now know that the basic unit of charge is actually $e/3$, which is the charge on the sub-subatomic particle called quarks that make up protons and neutrons. Quarks carry charges of $1/3e$ and $2/3e$, positive or negative.

- Electrons carry charges of $-e$, and protons carry charges of $+e$, even though those particles are dramatically different from one another.

- The SI unit of charge is the coulomb (C), , which is named after Charles Coulomb, and 1 C is about $6 \times 10^{18}e$, which makes the elementary charge 1.6×10^{-19} C.

- The force between 2 electric charges depends on the 2 charges. If you have 2 objects that are completely uncharged, then they don't experience any electric force.

- If 2 objects have zero net charge, they may or may not experience an electric force, depending on whether the net charge is made up of several positive and negative charges.

- The farther apart 2 charges are, the weaker the force becomes. The force depends on the inverse square of the distance between them, and the direction is such that likes repel and opposites attract.

- Mathematically, the electric force is described by **Coulomb's law**, which states that the force between 2 charges is proportional to the product of the 2 charges and inversely proportional to the square of the distance between them: $F = \dfrac{kq_1q_2}{r^2}$.

- The force (F) between 2 charges, q_1 and q_2, is negative if the force is attractive, and positive if it's repulsive. The square of the distance between the charges is r. In the SI system, the constant k has the value 9×10^9 newton-meters squared per coulombs squared (N·m²/C²).

- The force between 2 charges can be compared to the force of gravity: They're both proportional to the inverse square of the distance between the objects.

- However, because there is only one kind of mass, gravity is always attractive. There are 2 kinds of charges, and the electric force can be attractive or repulsive. In addition, electric force is infinitely stronger than gravitational force.

- Quantitatively, the gravitational force between 2 protons is smaller than the electric force by a factor of 10^{36}, which is an enormous number.

- Even though Earth has a lot of electric charge in it, it has nearly a zero net electric charge, so it has no large-scale electrical effects.

- On the other hand, there's only one kind of mass with gravity, and that mass is only attractive, so large accumulations of mass come together. Even though gravity is the weakest of the fundamental forces, Earth's gravity—the gravity of large accumulations of matter—becomes quite strong.

Electric power lines are distributions of charge.

© iStockphoto/Thinkstock.

- Coulomb's law is only true when applied to the force between 2 tiny infinitesimal points of charge, like electrons and protons. When you begin to analyze objects that have complicated shapes, you have to look at the interactions of all the charges that make up those objects.

- Charged objects are called charge distributions. There are simple charge distributions; for example, the hydrogen atom consists of a proton and an electron. More common are complicated distributions of charge.

- Suppose there are 2 electric charges, q_1 and q_2, and we want to know what force they exert on another charge, q_3. Because electric forces add vectorially, the electric force from charge q_1 added to the electric force from charge q_2 gives the net electric force on q_3. This phenomenon is called the **superposition principle**.

- The force that charge q_2 exerts on charge q_3 is attractive because q_2 is negative and q_3 is positive.

- Using Coulomb's law, we can then calculate the force of charge q_1 on charge q_3 and the force on q_3 from q_2. We'd want to know this because we want to understand how charge q_3 is going to respond in the vicinity of charges q_1 and q_2.

- Applying the superposition principle shows that, from far away, a complicated charge distribution with some net charge produces essentially the same electric force as a single point charge—provided that the charge is nonzero and that the distribution is finite. This allows us to find the electrical influence of distributions of electric charge.

Important Terms

Coulomb's law: An equation that predicts the force between any 2 stationary charges at a given distance: $F = \dfrac{kq_1q_2}{r^2}$.

electric charge: The conserved quantity that acts as a source for the electric field.

electromagnetism: One of the 4 fundamental forces of nature that involves the interaction of particles having the property of charge; like charges repel, and unlike charges attract. Electromagnetic forces govern the behavior of matter from the scale of the atom to the scale of mountains.

superposition principle: This principle describes the phenomenon that electric forces add vectorially.

Suggested Reading

Rex and Wolfson, *ECP*, chap 15.1–15.3.

Wolfson, *EUP*, chap 20.1–20.2.

Questions to Consider

1. An electron and its antiparticle, the positron, collide and annihilate in a burst of pure energy. Because the electron and positron are charged, how does this not violate charge conservation?

2. The electric force is vastly stronger than gravity, yet gravity seems to be the more dominant force in our everyday world. Explain this apparent contradiction.

The Electric Field
Lecture 29

Gravity and the electric force both act between distant objects. An alternate description uses the concept of field. In this view, every mass creates a gravitational field in its vicinity, and other masses respond to the field right where they're located. Similarly, charges produce electric fields, and other charges respond to those fields. We represent electric fields with field lines, whose direction is that of the field and whose density represents the field strength. A charge placed in a field experiences a force proportional to its charge and the field strength.

- People used to think that Earth somehow reached out across empty space to pull on the Moon and that the 2 exerted forces on each other from a distance.

- In the field view, Earth creates a gravitational field that extends throughout space, and the Moon responds to what is going on in its immediate, local vicinity.

- Earth creates a gravitational field in the space all around it that reaches out and pulls an object with a force of 9.8 N/kg. The force on an object is its mass multiplied by the strength of Earth's gravitational field. Both the field and the force have direction—they're vectors.

- In terms of electrical charges, instead of thinking that a negative charge exerts a force on a distant positive charge, the negative charge creates a field in the space around it, and the positive charge responds. Equivalently, the positive charge creates a field in all the space around it, and the negative charge responds to the field in its immediate vicinity.

- The strength of the field is measured in newtons per coulombs, or N/C. The electric force on a charge q in a field E is therefore qE. The electric force is simply the charge times the electric field it experiences, just like the gravitational force was the mass times the field it was responding to.

- The simplest kind of system we know of electrically is a point charge. A point charge q_1 and a point charge q are some distance r apart: $F = \dfrac{kq_1q_2}{r^2}$.

- Instead of using this problem about a force between 2 specific charges, we want to consider the influence q creates in space all around it, particularly at this point q. Therefore, we replace the force q with a field: $E = \dfrac{kq}{r^2}$.

- If we happen to put a charge at that point, it will experience a force due to the field at that point, but the field exists either way.

- Mathematically, the **electric field** is the force per unit charge at q's location, $E = \dfrac{kq}{r^2}$, just as the gravitational field was the force per unit mass.

- This equation is just like Coulomb's law, except it doesn't have that second charge because the second charge is sort of latent. We can put anything we want there, any other charge, and the equation would give us the force on that charge.

- If you want to picture an electric field in 3 dimensions, you can imagine vectors at every point representing the field—that is, the direction and magnitude of the force per unit charge that might exist at any point.

- Starting at some arbitrary point, you can figure out which way the electric field points and draw **field lines** that begin at the point charge and extend radially outward, in principle, all the way to infinity.

- The field lines expand outward in 3 dimensions, and the density of field lines, the number of field lines that are crossing a given area, is becoming smaller and smaller—they are becoming farther apart.

- The electric field isn't flowing or moving, but these field lines describe the direction and the magnitude through their density of the electric field.

The plasma in the Sun's corona is composed of hot, ionized gasses, which are conductors because they have free charges.

- A more complicated charge distribution is called an **electric dipole**, which is composed of 2 point charges of equal magnitude but opposite signs. These 2 charges, one positive and one negative, make up a system that has zero net charge—yet the 2 charges are slightly separated.

- Dipoles are a metaphor for many molecules; if you understand dipoles, you understand much of the behavior of molecules.

- The electric field pattern of a dipole begins as a single field line that starts from the positive charge and radiates outward in an arch-like structure to the negative charge. There are infinitely many of these types of field lines.

- The dipole field at large distances actually falls off as $1/r^3$, faster than any point-charge field because the dipole has no net charge.

- **Gauss's law** states that the number of field lines that emerge from any closed surface depends only on the enclosed charge—a law that is exactly equivalent to Coulomb's law.

- Gauss's law provides a test of the inverse square law for the electric field. It's the reason the field lines spread out in space the way they do, and the reason that there aren't any new field lines except where there is charge.

- Mathematically, we use a calculus equation to describe Gauss's law: $\oint \mathbf{E} \cdot d\mathbf{A} = 4\pi k q_{enclosed}$.

- The expression on the left-hand side of the equation for Gauss's law is a mathematical surrogate for the idea of number of field lines. Overall, the equation states that the number of field lines is proportional to the enclosed charge.

- An **electrical conductor** is a material in which charges are free to move. Metals, for example, are conductors.

- Electric fields, of course, exert forces on charges. By definition, equilibrium is letting all charges move around as they want to until they stop. If a conductor is in equilibrium, there is no electric field inside. If there were, charges would move.

- You can do a Gaussian proof of this statement. If there is a conductor in equilibrium, any excess charge must lie on the surface of that conductor because, by definition, if a conductor is in equilibrium, charges are not moving.

- On a large scale, there are no electric fields in a conductor in equilibrium. The electric field is zero everywhere inside the conductor, so there are no field lines emerging from that surface, which is called a Gaussian surface.

Questions to Consider

1. Is the force on a charged particle always in the direction of the electric field at its location? If so, why? If not, give a counterexample.

2. Can 2 electric field lines cross? Explain.

3. An electric dipole has zero net charge. So why does it produce an electric field at all, and how does its field differ from that of a point charge?

Electric Potential

Lecture 30

M oving charges around in electric fields involves doing work against the electric force. Like gravity, the electric force is conservative—meaning that work done against it gets stored as potential energy and that the work involved in moving between 2 points is independent of the path taken. Therefore, the electric potential difference between 2 points is the work, or energy per unit charge, involved in moving between those points. Potential difference is essentially synonymous with voltage, and we can map electric potential differences with equipotentials, which are analogous to contour lines on a map.

- The electrical concept of voltage is very closely related to a concept called electric potential difference, a profound concept that links to the ideas of electric forces and electric fields.

- For the purposes of this course, with one exception, "voltage" and "electric potential difference" are essentially synonymous.

- Work and energy, as indicated by the work-energy theorem, are very closely related concepts. We will often use them essentially interchangeably and say that electric potential difference is about the movement of charge in electric fields and how much energy is involved in that motion.

- **Electric potential difference** is the work per unit charge needed to move charge between 2 points: $\Delta V_{AB} = E\Delta x$.

- Between 2 points, A and B, ΔV is the electric potential difference. Because potential difference is work per unit charge, electric charge is force per unit charge. Force times distance is work, so field times distance is work per charge.

- In mechanics, we learned that work was the area under the force-versus-position curve. Consequently, the electric potential difference is the area under the field-versus-position curve.

- Work is measured in joules, and charge is measured in coulombs, so the unit of electric potential difference (ΔV_{AB}) is J/C, energy per charge.

- Voltage is a measure of the energy involved in moving electric charge between 2 points, and it is the energy per unit of charge involved in moving between those 2 points. The unit J/C defines the volt.

The voltage of a battery can be quantified by a voltmeter, a device that measures many electrical quantities.

- For a 9-volt battery, the 9 volts tells you that if you move charge from one terminal of the battery to the other, you will either have to do 9 joules of work, or 9 joules of work will be done on you for every coulomb of charge you move between those terminals.

- We were measuring electric fields as N/C. In terms of volts as J/C, you can also write N/C as volts per meter. Therefore, volts per meter of distance is another measure of electric field strength.

- Like gravity, electric forces due to electric charges are conservative. We found out with conservative forces that it doesn't matter what paths you take to get from point A to point B—the amount of work had to be the same.

- We learned that the electric field of a point charge is a field that projects radially outward and falls off as $1/r^2$.

- Similar to the gravitational potential energy associated with moving around in the inverse square gravitational field, the potential difference from point A to point B in the electric field of a point charge is $kq/r_B - kq/r_A$.

- The potential in the field of a point charge, $V(r)$, is simply kq/r for a point charge: $V(r) = kq/r$. The potential difference between a point infinitely far away and a point a distance r from a point charge is this quantity kq/r.

- A flat map represents the third dimension with **contour lines**, which are lines of constant elevation. You could walk along a contour line on the side of a mountain, and you wouldn't be going up or down, so you wouldn't be doing any work against gravity.

- **Equipotentials** are at right angles to the electric field, just as contour lines on a map are at right angles to the steepest slope.

- Equipotentials have to be at right angles to the electric field because the only way you can do no work while moving against a field is to not move against it—to be perpendicular to it. Equipotentials can be represented by contour lines.

- On a map containing equipotentials, like the one in Figure 30.1, the electric field is strongest where the potential changes most rapidly, just as terrain is steepest where the contour lines are closest and the height changes most rapidly.

- In Figure 30.1, the electric field is pointing to the left, and the maximum electric field is where the contours are closest together, which is marked by $\mathbf{E}_{max}$ in the figure.

- The labels on the contours are measures of potential, and they're all relative to some area where the potential is defined to be zero.

Figure 30.1
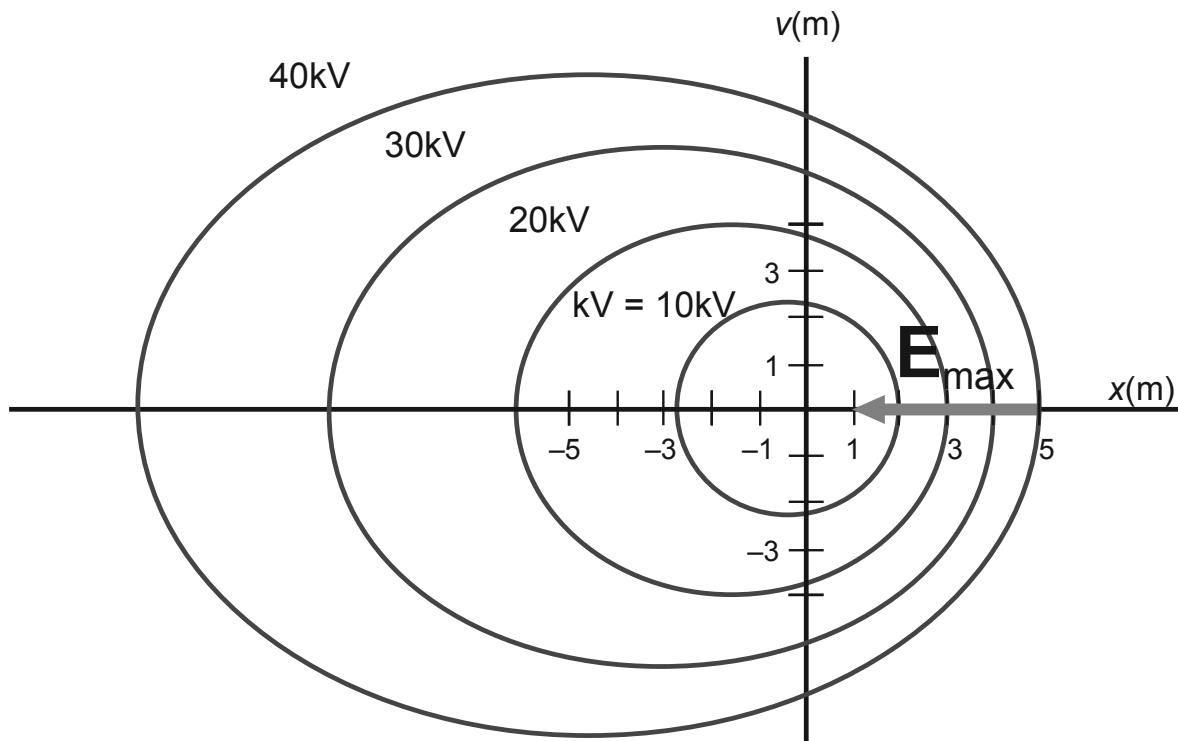
- To find the greatest electric field, we use the equation $\mathbf{E}_{max} = kq/r_B - kq/r_A$. We start at 5 m, where the potential is 40 kV, and travel to 2 m, where the potential is 10 kV: (40 kV − 10 kV)/(5 m − 2 m) = 30 kV/3 m = 10 kV/m. A volt per meter is equivalent to a newton per coulomb, so the greatest electric field occurs at 10 kN/C, to the left.

- What's the force if you put a 1-mC charge in that field? Force is the charge times the electric field strength, so that's 1 mC times 10 kN/C, which comes out to 10 N by cancelling out the units.

- Suppose you have an electrical conductor, which has no electric field within it, and you consider the electric field right at its surface.

- If the conductor is in equilibrium, the field must be perpendicular to the surface because charge can't move around on the surface. If the field had a component parallel to the surface, charge would move, and it wouldn't be in equilibrium.

- Since the equipotentials are at right angles to the field, equipotentials right near the surface have to be basically parallel to the surface.

- If you have a charge conductor, some irregularly shaped object, and you draw equipotentials around it, the ones closest to the object are going to hug its surface. The ones farther out will hug the surface less.

- By drawing field lines, you find that they have to concentrate in the regions where the equipotentials are hugging around some pointed part of the irregularly shaped object. This means that the electric field of a charged object is going to be strongest where the object curves most sharply.

- The application of this idea is if you're designing systems that operate at high voltage, you need to avoid sharp corners.

- For example, a corona discharge on a power line is a place where there are nuts or bolts with sharp edges. At these sharp edges on a high-voltage power line, there are strong enough electric fields that break down the air so that sparks are jumping off the power line, which causes the loss of electrical energy and a dangerous situation.

Important Terms

contour line: A line of constant elevation on a map that is perpendicular to the steepest slope.

electric potential difference: The work per unit charge needed to move charge between 2 points, *A* and *B*: $\Delta V_{AB} = E\Delta x$.

equipotential: Just as contour lines on a map are at right angles to the steepest slope, these lines are at right angles to the electric field.

Lecture 30: Electric Potential

Suggested Reading

Rex and Wolfson, *ECP*, chap 16.1–16.3.

Wolfson, *EUP*, chap 22.

Questions to Consider

1. What is a volt?

2. A bird lands on a 7200-V power line. Why doesn't it get electrocuted?

3. If the electric field is zero at some point, does that tell you anything about the electric potential at that point?

Electric Energy

Lecture 31

$\mathbf{B}$ringing widely separated charges together involves work, and the resulting charge distribution has stored potential energy. With 2 positive or 2 negative charges, the stored potential energy is positive and could be recovered by releasing the charges. With opposite charges, the stored energy is negative, and it would take energy to separate the charges. A technological system for storing electric energy is the capacitor, consisting of a pair of separated electrical conductors. Analyzing energy in capacitors reveals a universal fact: All electric fields represent stored energy.

- Every arrangement of electric charge—whether it's something we build technologically or something nature builds in a molecule—represents stored electrostatic energy.

- Starting with a point charge q_1, we can imagine bringing a second point charge, q_2, from very far away to the vicinity of point charge q_1. It ends up a distance r from point charge q_1, and the potential at point r is the work per unit charge to come in from infinity to that point, which is kq_1/r.

- Because the electric force is conservative, the work we did ends up being stored as electrostatic potential energy. The stored potential energy of a system of 2 point charges is kq_1q_2/r.

- In this case, that stored energy is positive because we had to do work to bring those 2 charges together.

- In the same situation, except q_2 is a negative charge, q_1 is going to experience an attractive force, and work is going to be done on that charge.

- Equally, we would have had to hold the charge back with a force going the opposite way if we didn't want q_2 to move toward the positive charge.

- Consequently, the stored potential energy is now negative, which means that if you wanted to remove the 2 charges to a very large distance, you would have to do work that was equal to the negative energy but a positive amount of work of the same magnitude.

- Negative potential energies generally describe a system in which 2 objects—for example, molecules—are bound together by an attractive force.

- With some simple calculations, we are able to estimate the energy that is released when molecules are formed or, equivalently, the energy needed to take apart molecules.

- In technological applications, a **capacitor** is our primary energy-storage device, and it consists of a pair of electrical conductors whose charges are equal but opposite.

- A **parallel-plate capacitor** contains a pair of parallel conducting plates that are broad in area, as compared to the relatively narrow spacing between them.

- **Capacitance** is the measure of how much charge a capacitor can hold.

- A pair of metal plates in a parallel-plate capacitor has area A, and the spacing between the plates is d. The charge $+q$ is on one plate, and the charge $-q$ is on the other plate.

- Although a capacitor remains electrically neutral, you can put a charge on one conducting plate and the opposite charge on the other. As a result, an electric field develops in the region between them, and if the plates are parallel and quite closely spaced, that field will be essentially uniform.

- There will be, as a result of that electric field, a potential difference (V) because you have to do work moving against the field.

- Capacitance is the ratio of the charge $+Q$ or $-Q$ divided by the potential difference (or voltage): $C = Q/V$.

- This ratio has the units of coulombs (charge) per volt (voltage), and that unit defines what's called a farad, and 1 farad is a very large capacitance.

- The more charge you can put between the plates for a given voltage, the bigger its capacitance.

- The bigger the area of the plates, the more charge you can put on them for a given voltage. Consequently, the capacitance depends on the area; it scales linearly with the area—if you double the area, you double the capacitance.

- On the other hand, if you separate the plates for a given charge, the electric field stays essentially the same, but it's over a bigger distance, so the voltage is bigger, and the capacitance (Q/V) decreases.

A computer video card contains memory chips that store display data. A 1-gigabyte memory chip could have about 8 billion capacitors.

- The capacitance is inversely proportional to the separation distance (d), and the capacitance scales as A/d with a constant, which involves the coulomb constant: $C = \dfrac{A}{4\pi kd}$.

- Often we put an insulating material between the plates of a parallel-plate capacitor—to keep the plates apart and, more importantly, to change the capacitance.

- **Insulating materials** don't conduct, but they consist of little dipoles, or molecules. The dipoles flip and align themselves with the electric field, so the molecules act as dipoles in insulating material.

- When the dipoles align, the dipole field is opposite the field that was applied because the negative part of the dipole is attracted to the positive plate.

- As a result, this effect reduces the overall electric field in the capacitor. That, in turn, lowers the voltage for a given charge, which leads to a greater capacitance. Real capacitors have insulation, which serves to increase the capacitance.

- Capacitors not only store electric charge, they also store energy because they are systems of charges—electrostatic systems—so energy is involved.

- The energy stored in a capacitor is equal to 1/2 times the capacitance times the squared voltage: $U = 1/2(CV^2)$.

- When the voltage doubles, the energy is multiplied by 4, so that means high-voltage capacitors are very good energy storage systems.

- Capacitors can store energy, and then they can release that energy very quickly.

- Using the equation for energy stored in a capacitor, $U = 1/2(CV^2)$, and the voltage between plates, $V = Ed$, and the capacitance, $C = Q/V = A/(4\pi kd)$, we can derive the following:
$$U = 1/2(CV^2) = \frac{1}{2}\frac{A}{4\pi kd}E^2 d^2 = \frac{Ad}{8\pi k}E^2 = \frac{1}{8\pi k}E^2(Ad).$$

- The energy stored in a capacitor is a constant times the square of the electric field times the separation of the plates multiplied by their area, which is the volume inside the capacitor (area of a rectangular region times the height of that rectangular region): $\frac{1}{8\pi k} E^2 Ad$.

- This resulting quantity is the density at which energy is stored in an electric field. We can multiply that density, joules per cubic meter (energy per volume), by the volume, and we get the total energy stored in the capacitor.

- Whenever there is an electric field anywhere in the universe, it represents stored energy—the stored energy density at the point where the electric field has strength E and is proportional to E^2.

- Therefore, if you double the electric field, you quadruple the energy density. This is a profound statement because it tells us that electric fields have substance to them.

- We cannot use capacitors as the primary energy-storage motive for cars, especially air-insulated capacitors, because they are so different from anything like the energy density of gasoline.

- We can temporarily store energy in capacitors for trains, for example: The kinetic energy the train has as it is slowing down gets stored as energy in capacitors and then is used again as the train starts up. However, most of a train's motive comes from power plants.

Important Terms

capacitance: The measure of how much charge a capacitor can hold.

capacitor: An energy-storage device that consists of a pair of electrical conductors whose charges are equal but opposite.

insulating material: A material with no or few free electric charges and, thus, a poor carrier of electric current.

parallel-plate capacitor: A capacitor that contains a pair of parallel conducting plates that are broad in area, as compared to the relatively narrow spacing between them.

Suggested Reading

Rex and Wolfson, *ECP*, chap 16.4, 16.5.

Wolfson, *EUP*, chap 23.

Questions to Consider

1. A dipole consists of 2 opposite charges of equal magnitude. Is the electrostatic energy stored in the dipole's electric field positive, negative, or zero?

2. Does a capacitor of a given capacitance have the ability to hold a certain fixed amount of charge in the same way a 1-gallon bucket can hold 1 gallon of liquid?

3. Explain the role of the capacitor in your camera's flash circuitry.

Electric Current
Lecture 32

The flow of charge in electrical conductors constitutes electric current. A flow of 1 coulomb of charge every second is a current of 1 ampere. In many conductors, the current is proportional to the electric potential difference across the conductor and inversely proportional to the conductor's electrical resistance—this is Ohm's law. Multiplying voltage (energy per charge) by current (charge per time) gives energy per time, or electric power. Electricity is a versatile and convenient form of energy, but it can be dangerous.

- Electric current is what flows in wires, and it's what you think of when you think of the word electricity—the flow of electrons in wires to lightbulbs and electric stoves.

- Because we are now allowing charge to move, we're no longer dealing with electrostatic situations. When we had conductors charged up, or capacitors fully charged, we had electrostatic equilibrium.

- We're now going to abandon the assumption of equilibrium, and the electric field is no longer zero in a conductor that isn't in electrostatic equilibrium.

- **Electric current** (*I*) is a flow of charge—it's the net rate of charge crossing a given area.

- The unit of current—because it's charge per time—is coulombs per second, which is known in the SI system as an ampere (A), or amp, and is named after the French scientist Ampère. A typical incandescent lightbulb of 75 or 100 watts draws a current of about 1 ampere.

- Currents can run through whatever area you specify, but usually we're talking about conductors whose area is easy to define—like a wire.

- In order for there to be a current in a material, there has to be a net flow of charge in one direction or the other—either positive or negative charges have to be moving while the others are stationary, or both could be moving in opposite directions.

- If both charges are moving equally in the same direction, there's no net current. Just because there are charges in a material and the material is moving does not mean there is a current.

- In metals, the current is carried by free electrons. In ionic solutions—sports drinks, for example—the current is carried by both positive and negative ions. Plasmas, like ionic solutions, carry current both ways.

- **Semiconductors**, the devices that make modern electronics work, are engineered to have particular concentrations of electrons and holes, which are simply the absence of an electron. They move through the material as if they were a positive charge but have both types of charge.

- **Superconductors** are devices that conduct electricity with no loss of energy; they are perfect conductors. Superconductors require very cold temperatures to operate.

- In superconductors, a complicated mechanism involving quantum mechanics causes widely separated electrons to pair up and then move as waves with no resistance, no loss of energy, through the conductor.

- Metals are like most solids, which if we let them form naturally and cool down slowly, tend to form regular arrays of atoms called crystals. The structure of a crystal is called a lattice, which is regular geometrical spacing.

- In metals, the outermost electrons of individual atoms are so loosely bound to their atoms that when the whole structure forms into a crystal, the electrons become freed and are able to wander through the metal without being attached to individual atoms.

- There is a sea of free electrons surrounding the positive ions, and at ordinary temperatures, these free electrons move around randomly at high speeds.

- A single electron is in random thermal motion. When it runs into ions, it typically gives up some of its energy, but it also may gain energy back from the ions.

- On average, electrons don't go anywhere. There's no current, and there's no net flow of electrons.

- In the presence of an electric field in a metallic conductor, electrons acquire a slow drift velocity, which occurs because electrons are negative—opposite the direction of the field.

- Drift velocity is proportional to the electric field. The bigger the electric field, the bigger the small amount of change that develops in the rapid, random thermal motion of the electrons.

- The electrons are negative, so there's a current in the direction opposite the field. Current is proportional to drift velocity but with a negative sign—it's proportional to the opposite of the electric field.

- The voltage across the conductor is also proportional to the electric field because voltage is electric field times distance, so the current is proportional to the voltage in an electrical conductor like a metal.

- Ohm's law is an approximate law that describes how some materials, particularly metals, behave. It's empirical—something we observe in the world.

- **Ohm's law** is a relationship between electric voltage (V), current (I), and **resistance** (R), which is a measure of how much a conductor impedes the flow of currents. In many conductors, current is proportional to voltage: $V = IR$.

A typical incandescent lightbulb of 75 or 100 watts draws a current of about 1 ampere.

- Resistance depends on the material properties of the conductor, such as the density of electrons with it, the crystal structure, and the presence of impurities.

- Resistance also depends on temperature; usually, as the temperature increases for metals, the resistance also increases. In addition, resistance depends on the size and shape of the conductor.

- Voltage is energy per charge, and it tries to push a current through a resistance. The bigger the voltage, the more current you'll get. The bigger the resistance, the less current you'll get.

- The unit of resistance is the ohm (Ω), which is named after German physicist Georg Simon Ohm and is a measure of volts per amp.

- Energy per time is a measure of power; multiplying voltage and current, you end up with power in watts (as used in the SI system).

- **Electric power** is $P = IV$, and there are 2 alternate forms that can be derived from this equation: If you take $V = IR$, you get $P = I^2R$. If you take $I = V/R$, you get $P = V^2/R$.

- If you increase I, then you also increase the power loss, and that is the reason large currents aren't used in power lines. To supply the same power while keeping I small, the voltage has to be very large, which is why power is transmitted at very high voltages in power lines.

- Current passing through your body is dangerous, and it would take a very high voltage to drive significant current through your body under normal, dry conditions. It's neither high voltage nor high current alone that's dangerous—it's a combination.

- To keep ourselves safe in the electrical context, Earth's ground provides a zero of potential, and one side of an electric power line is grounded.

- Many devices that are likely to be used in situations where you would be in contact with grounded objects—like water pipes, for example, in the kitchen when using power tools—have a third wire, a ground wire.

- Then, if there's a short circuit, the current flows down the ground wire and breaks the circuit breaker or fuse that protects the circuit from short circuits. Otherwise, a fire could result because too much current would be flowing.

- Devices called ground fault circuit interrupters sense the current flowing down one wire and back through the other wire and shut off the circuit before a dangerous current can develop.

Important Terms

electric current: A net flow of electric charge.

electric power: The rate of producing or expending energy. In electrical devices, power is the product of voltage and current.

Ohm's law: The statement, valid for some materials, that the electric current is proportional to the applied voltage and inversely proportional to the material's resistance.

resistance: The property of a material that describes how it impedes the flow of electric current.

semiconductor: A material that lies between insulators and conductors in its capacity to carry electric current. The electrical properties of semiconductors are readily manipulated to make the myriad devices at the heart of modern electronics.

superconductor: A material that, at sufficiently low temperature, exhibits zero resistance to the flow of electric current.

Suggested Reading

Rex and Wolfson, *ECP*, chap 17.1–17.4.

Wolfson, *EUP*, chap 24.

Questions to Consider

1. All matter contains electric charges in the form of electrons and protons. So, if you walk along carrying a piece of matter, does that constitute an electric current? Why or why not?

2. The electric current in a metallic conductor is proportional to the electric field in the conductor, implying that the electrons have a constant drift velocity proportional to the field strength—yet Newton's second law suggests that the field should give the electrons a constant acceleration. Explain why this is not a contradiction.

3. Explain why many electrical devices have 3-wire grounding plugs.

Electric Circuits
Lecture 33

E lectric circuits—interconnections of electrical components—require an energy source, such as a battery, an electric generator, or a solar cell. In a battery, chemical reactions separate positive and negative charges to the battery's 2 terminals. Connect an external circuit, and current flows from one terminal to the other. In the process, the battery supplies energy to the external circuit. There are 2 ways to connect electrical components: series and parallel. When capacitors are incorporated into circuits, they store electric energy and introduce time dependence into the circuit's behavior.

- **Electric circuits** are assemblages of electronic components and devices, and they almost always have a source of electrical energy. They are ubiquitous: They're in all of our electronics and appliances—even houses are electric circuits.

- An **emf**, which stands for electromotive force, is a source of electrical energy that maintains a constant potential difference across 2 electrical terminals in a circuit.

- Batteries are sources of emf, and so are electric generators and power supplies, which are plugged in to produce a fixed voltage across their 2 terminals.

- The simplest possible circuit includes a source of emf and an electrical load—for example, a battery and a lightbulb, respectively.

- Charge flows through that circuit from the battery's positive terminal through wires—through the external load—to the negative terminal.

- It makes no difference whether there is a positive charge moving from the positive to the negative or a negative charge moving from the negative to the positive—the current is still moving in the same direction (from positive to negative).

- When the positive charge arrives at the negative terminal, it experiences an electric field inside the battery that points downward from positive to negative.

- Then, chemical reactions inside the battery do work on that charge, lifting it against the battery's internal electrical field to deliver it back to the positive terminal.

- The battery does this work at the expense of its internal chemical energy, which is why a battery runs down and either has to be discarded or recharged.

- A series combination is a slightly more complicated circuit consisting of a single battery and 2 loads (lightbulbs) that are connected one after the other.

- In a series combination, charge flows from the positive terminal through the load and back around to the negative terminal, and once again, it's boosted inside the battery. Whatever energy the charge gains from the battery, it gives up in the load.

© Hemera Technologies/PhotoObjects.net/
© Getty Images/Thinkstock.

Power supplies, which are used to generate electricity, are plugged in to produce a fixed voltage.

- In circuits, the battery has some voltage, V_b, that it is supposed to maintain across its terminals. There's some voltage across the first load and some voltage across the second load, and those voltages represent the energy per unit charge that's lost as charge moves through those loads.

- In this case, the loads are lightbulbs, so that energy is converted into visible light and heat in the lightbulb.

- In order for this simple circuit to conserve energy, the sum of the 2 voltages across the load has to sum to the battery voltage. In other systems, there may be more load voltages that also have to sum to the battery voltage.

- There are 2 ways to connect electrical components, such as **resistors**—devices engineered to have specific resistance— capacitors, and batteries: in series and in parallel.

- If components are in series, the only place current can go after it moves through one component is through the next component. In a series circuit, the flow of current in a steady state is the same in all elements.

- In a parallel circuit, the resistors are connected across the battery. Because they're connected across the battery and the battery maintains a fixed voltage or potential difference across its terminals, the resistors get the same voltage.

- In a parallel circuit, the resistors have the same voltage, but the currents flowing through the resistors have to add up to the current that comes from the battery (or else charge would not be conserved).

- Figure 33.1 is a series resistor circuit, which includes a battery with voltage V_b, resistor R_1, and resistor R_2.

- Energy conservation shows that the sum of the voltages across the 2 resistors have to add up to the battery voltage: $V_1 + V_2 = V_b$.

- In series resistors, the same current is flowing through each resistor because once charge flows out of the battery through R_1, there's no place for it to go but through R_2.

Figure 33.1
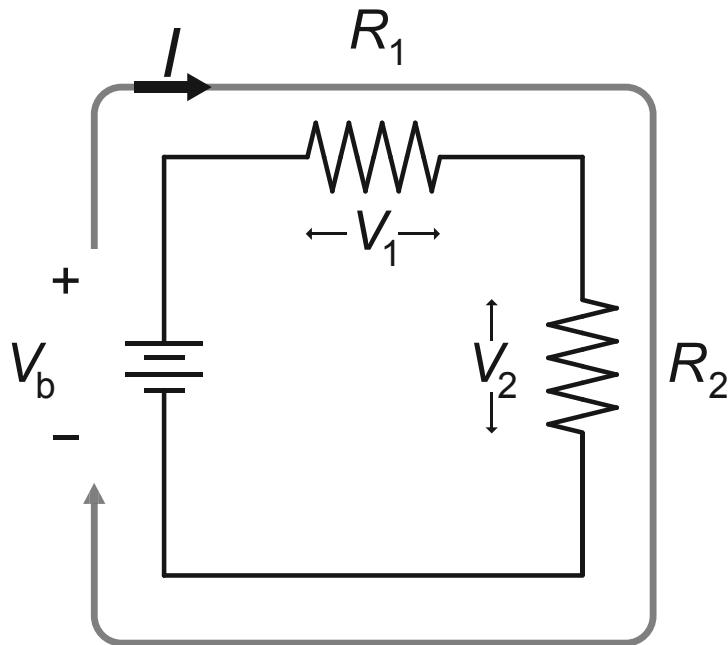
- Ohm's law says that voltage is the product of current and resistance. Therefore, the voltage across the first resistor is $V_1 = IR_1$, and $V_2 = IR_2$.

- Using these 2 equations in the statement of energy conservation, $IR_1 + IR_2 = V_b$. Algebraically, $I = V_b/(R_1 + R_2)$.

- This looks just like Ohm's law, $I = V/R$, except that instead of R, we have the sum $R_1 + R_2$. The combination of these 2 resistors adds up to a series resistance, which is the sum of the 2 resistances: $R_{series} = R_1 + R_2$.

- Series resistors add, and this argument can extend to multiple resistors in a series, which are equivalent to a single resistor whose resistance is their sum.

- Series resistors divide the voltage of the battery—or whatever the source is—into 2 parts, and those 2 parts divide, in proportion, to the 2 resistors.

- The individual voltages are $V_1 = IR_1 = (V_b/(R_1 + R_2))R_1$ and $V_2 = IR_2 = (V_b/(R_1 + R_2))R_2$. The quantity $R_1/(R_1 + R_2)$ for V_1 and $R_2/(R_1 + R_2)$ for V_2 is called the voltage divider fraction.

- This circuit is a voltage divider, which divides the voltage from the battery into 2 pieces that are not necessarily equal: The bigger the resistor has the bigger voltage across it.

- An ideal battery is impossible. If it existed, it would be the solution to all our energy problems because it could supply infinite energy.

- A real battery looks like an ideal battery with a resistor in series with it. The resistor represents the slowness of the chemical reactions that are going on inside the battery.

- Internal resistance is what distinguishes real batteries from ideal ones. If you want to use a battery and make it work well, you should use it in such a way that the internal resistance is low compared to the load you're connecting across it.

Figure 33.2
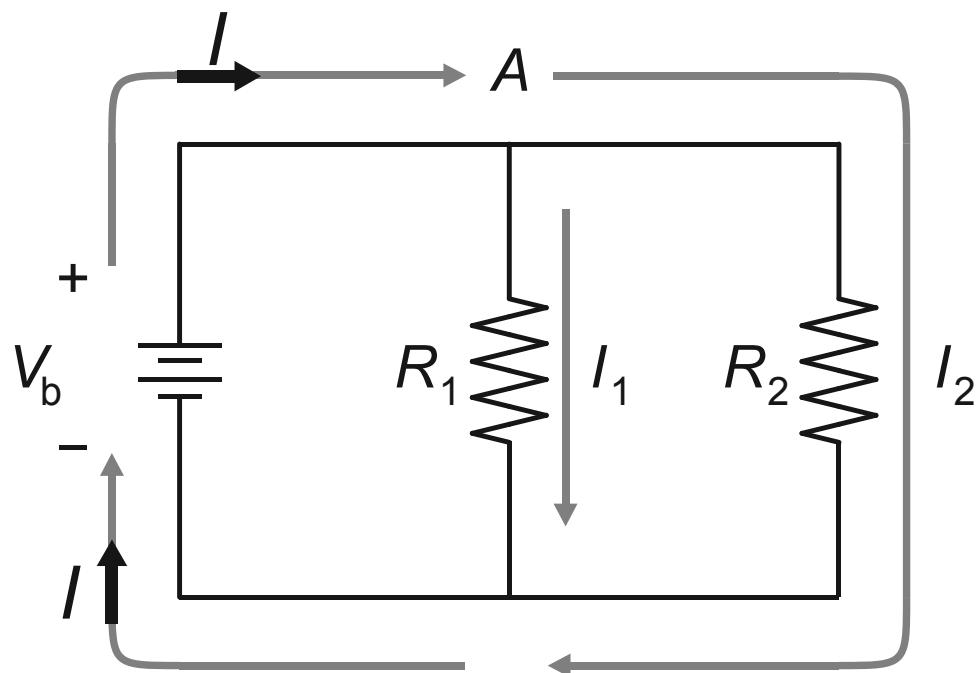
- Parallel resistors, like the one in Figure 33.2, have the same voltage across them because they're both connected across the battery.

- Charge has to be conserved, so the battery delivers a current I that is the sum of the current I_1 through resistor R_1 and I_2 through resistor R_2. They're not in series, so they don't have to have the same current.

- According to Ohm's law, $I = V_b/R_1 + V_b/R_2$. Using algebra, $1/R_{parallel} = 1/R_1 + 1/R_2$. Therefore, parallel resistors add reciprocally.

- More complicated circuit systems have several resistors in a parallel combination or in series.

- Capacitors fundamentally introduce a time delay into a circuit. Capacitors smooth out alternating voltages to make them steady and smooth because they introduce time delays.

Important Terms

electric circuit: An electrically conducting path that can carry current in a loop.

emf: A term that stands for electromotive force and is a source of electrical energy that maintains a constant potential difference across 2 electrical terminals in a circuit.

resistor: An element in a circuit formulated to have a specific electrical resistance; it reduces the current that can pass for a given voltage.

Suggested Reading

Rex and Wolfson, *ECP*, chap 17.4, 17.5.

Wolfson, *EUP*, chap 25.

Questions to Consider

1. An ideal battery would maintain a fixed potential difference across its terminals, regardless of what's connected across it. Why is an ideal battery impossible?

2. The combined resistance of 2 parallel resistors is lower than that of either resistor alone. Why does this make sense?

3. Give one example of a natural system that can be modeled as an electric circuit.

Lecture 33: Electric Circuits

Magnetism
Lecture 34

M agnetism is fundamentally about moving electric charges. Magnetic fields exert forces on moving charges, but the force is more complicated than in the electrical case because it involves not only the charge and the magnetic field, but also the charge's motion and its orientation relative to the field. As a result, charged particles undergo complex trajectories in magnetic fields. In the simplest case, those trajectories are circles or spirals. The electric motor works because of forces on current-carrying wires.

- **Magnets** have 2 poles: north and south. North poles repel each other and south poles repel each other, while north attracts south and south attracts north.

- A **magnetic field** exerts forces on moving charges and has field lines, similar to those of a dipole.

- A relatively small class of materials, including iron and steel, are affected by magnetism. This class of materials does not include other common metals like aluminum or copper.

- Most importantly, magnetism is about moving electric charge—electricity—and moving electric charge responds to magnetic fields.

- The magnetic force acts on moving charged particles; charged particles at rest don't feel magnetism.

- Moving charged particles experience a magnetic force that's proportional to the electric charge q and magnetic field strength $\mathbf{B}$: $\mathbf{F} = q\mathbf{v}\mathbf{B}$.

- Magnetic force is more complicated than electric force because it involves the direction of the velocity in relation to the field (direction and magnitude), which is why vectors are involved in the equation.

- The magnitude of magnetic force is $F = qvB\sin\theta$.

- The unit of magnetic field is newton-seconds per coulomb-meter (N·s/C·m), which is given the name tesla after Nikola Tesla, a Serbian-American inventor.

- The direction of the magnetic force is perpendicular to both the velocity and the magnetic field, which are both vectors.

- Figure 34.1 depicts cyclotron motion. Imagine that there is a magnetic field that is pointing out of the page of this 2-dimensional guidebook, which is shown by the dots in the figure.

Figure 34.1
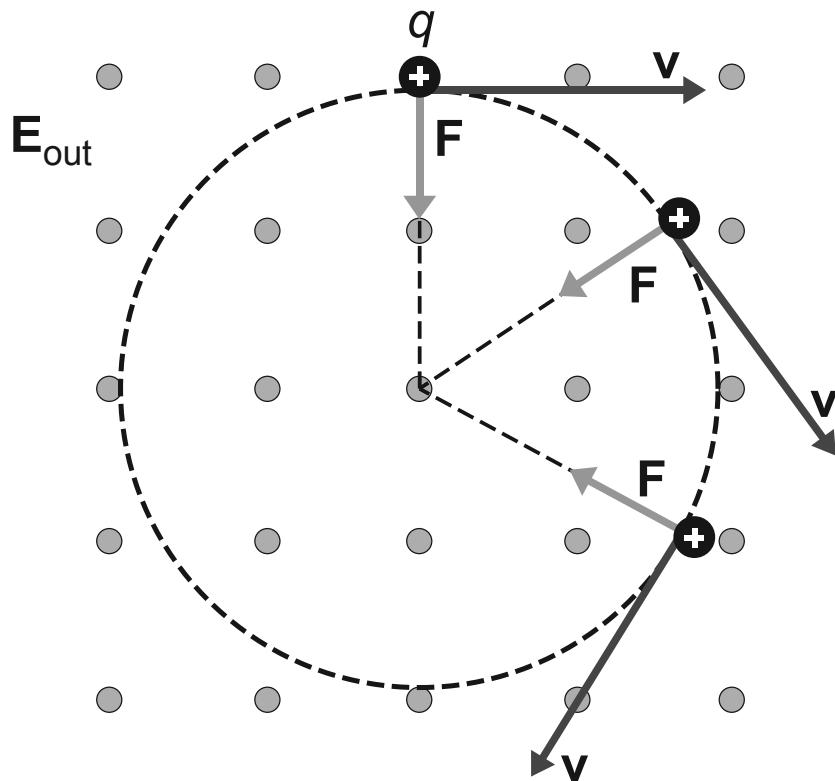
- In this figure, there is a charged particle with positive charge moving to the right with some velocity **v**. We have to rotate **v** onto **B**, which is pointing out, so the magnetic force is down.

- The direction of the magnetic force is perpendicular to both **v** and **B**, and we now know that it is pointing downward, so there is a force at right angles to the velocity.

- Forces that act at right angles to a velocity change only direction and not magnitude. Therefore, this velocity is going to have its direction changed, but its magnitude is going to be the same.

- The velocity continues to change in direction but not in speed as the charged particle travels around the circle. The result of that kind of motion of force produces uniform circular motion, so we do not need to use vectors.

- Applying Newton's second law ($F = ma$) and knowing that the acceleration in circular motion is $a = v^2/r$, we get $qvB = mv^2/r$. Algebraically, we solve for the radius: $r = mv/(qB)$.

- Velocity is distance over time, and $2\pi r$ is the circumference of this circle, so divide that by the period (T), the time it takes to travel around the circle, and you get $v = 2\pi r/T$. Then, we can rewrite $r = mv/(qB)$ as $r = m(2\pi r/T)/(qB)$.

- Using algebra, $T = 2\pi m/(qB)$, which is the cyclotron period. If we take its inverse, we get the cyclotron frequency: $f = 1/T = qB/(2\pi m)$. This is the number of circles it makes per unit time per second in the SI system.

- The cyclotron frequency and period are independent of the radius, the speed, and the particle's energy—the properties of the particle and the strength of the magnetic field.

- The idea of cyclotron motion only holds true for velocities and energies that are relatively low. The period and frequency of this motion are completely independent of the particle's energy.

- In a cyclotron particle accelerator, particles—protons and nuclei of various atoms—are accelerated in ever-enlarged spiral paths.

- The particles are given energy by an electric field, but they're held in place by a magnetic field, and as their paths spiral ever larger, the frequency—the number of turns they make in a unit time—doesn't change, which makes it very easy to keep the whole thing in synchronism.

- There are many applications of cyclotron motion, both in practical technological devices (microwaves) as well as in nature (astrophysical magnetic fields).

- In 3 dimensions, charged particles spiral rapidly around magnetic field lines. For example, high-energy charged particles from the Sun get trapped in Earth's magnetic field, causing auroras in the polar regions of the planet.

- The magnetic force on a current carrying wires involves both magnetic and electrical interactions—the upward force on the electrons and the magnetic force, which give the electric force on the ions.

- If you have a long, straight wire perpendicular to a magnetic field, the force is the current times the length of the wire times the strength of the magnetic field. If it's oriented at some other angle, you must include $\sin(\theta)$.

- It doesn't matter how fast the individual charge carriers are going or what their charge is—the magnetic force on a conducting wire depends only on the current, the length of the wire, and the magnetic field.

- It doesn't matter if you have a current loop in a magnetic field made of wire or an atomic current loop made of charged particles circling around in atoms: Current loops tend to align themselves with magnetic fields.

- A current loop has a **magnetic moment vector**, which is a vector perpendicular to the loop area, and its magnitude μ is the current in the loop times the area of the loop. This turns out to be true for any loop shape.

- The torque on the loop rotates the loop into alignment with the field and is the magnetic moment, which depends on the current, the area, the magnetic field strength, and $\sin(\theta)$, where θ is the angle that determines whether the system is out of or in alignment with the field.

© iStockphoto/Thinkstock.

Microwave ovens have a tube called a magnetron that generates the microwaves that cook your food.

- When the system lines up completely, the magnetic moment is aligned with the field. Then, there's no more torque, and the loop sits in its position.

- With a clever change in the direction of the current, we can keep the loop rotating, making an **electric motor**, which is basically a current loop that is placed in a magnetic field—typically between the poles of a magnet—and rotates on bearings on a shaft.

- When a (stationary) battery is connected to a circular loop (rotating), the way we get the current in the loop is we use small semicircular rotating copper (or other metal pieces) and brushes, which are typically made of wire or carbon.

- These pieces are in contact with the rotating surfaces, and they let electricity flow through the loop. This connection is called the commutator.

- Normally, the system aligns itself with the magnetic field, but just as it gets near alignment, the commutator switches past a gap within it that reverses which part of the battery is connected to which part of the loop, reversing the direction of the current in the loop.

- Then, the loop is no longer in alignment with the field, so it wants to rotate 180˚ to get back into alignment, but just as it barely gets there, the commutator reverses the direction of the current again, and the loop keeps rotating around. This is how an electric motor works.

Important Terms

electric motor: A current loop that is placed in a magnetic field—typically between the poles of a magnet—and rotates on bearings on a shaft.

magnet: An object that has 2 poles: north and south. North poles repel each other and south poles repel each other, while north attracts south and south attracts north.

magnetic field: The influence surrounding a moving electric charge (and, thus, a magnet) that results in forces on other moving charges (and on magnets or magnetic materials).

magnetic moment vector: A vector that is perpendicular to the area of a current loop. Its magnitude μ is the current in the loop times the area of the loop.

Suggested Reading

Rex and Wolfson, *ECP*, chap 18.1–18.3.

Wolfson, *EUP*, chap 26.1–26.4.

Lecture 34: Magnetism

Questions to Consider

1. Under what conditions will a charged particle in a magnetic field *not* experience a magnetic force?

2. A charged particle is undergoing circular motion perpendicular to a uniform magnetic field. If you increase the particle's speed, that doesn't change the time it takes to complete its circle. Why not?

3. What's the role of the commutator in a DC electric motor?

The Origin of Magnetism
Lecture 35

Moving electric charge is the source of magnetic fields. Although there might exist magnetic monopoles, none have ever been found, so moving electric charge is the only known source of magnetic fields. Common configurations include long, straight current-carrying wires and closed loops of electric current. Current loops produce magnetic fields that, from afar, have the same configuration as the electric field of an electric dipole. On atomic scales, magnetic fields arise from orbital and spinning motions of atomic electrons.

- It is impossible to split a magnet into a separate north and south pole; if you were to break a magnet into smaller pieces, each piece would be a complete magnet with both a north and a south pole. There is no such thing as an isolated magnetic pole, which would be called a **magnetic monopole**.

- Gauss's law states that if we draw a closed surface around some charge distribution, the total number of field lines emerging from that closed surface depends on nothing but the enclosed charge.

- There is no magnetic charge, so any closed surface will have zero magnetic field lines emerging from it. Magnetic field lines don't begin or end because when we had electric field lines, they began or ended only on charge.

- Magnetism is fundamentally about moving electric charge: Moving electric charge is both what responds to magnetism and is the source of magnetism.

- Let's look at how the magnetic field arises from an electric current. A piece of wire carries a current I, which is moving electric charge. There is some point p off to the right where you want to evaluate the magnetic field.

- The **Biot–Savart law** says that a very short length of current produces a magnetic field that falls off as $1/r^2$, just as the electric field of a point charge falls off: $\Delta B = \dfrac{\mu_0}{4\pi} \dfrac{I\Delta L r}{r^2}$.

- The field direction is perpendicular to both the current and a line extending from the current to the point p.

- Basically, this is the magnetic analog of Coulomb's law—the difference is that you can't have just a single, tiny, isolated piece of current. If you did, charge wouldn't be conserved.

- Although the magnetic field of an infinitesimal piece of the current falls of as $1/r^2$, when you add vectorially the magnetic fields of all the pieces of the current, you will never get a magnetic field that falls off as $1/r^2$.

- If you had a magnetic field that fell off as $1/r^2$, it would be the field of a magnetic monopole—just like the electric field that falls off as $1/r^2$ is the field of an electric dipole.

- The magnetic field of a current is given by the Biot–Savart law. What happens is the magnetic field lines ultimately wrap around the currents, which are their sources, to form closed loops.

- Magnetic field lines can't begin or end, so even though moving electric charge is the source of magnetic field, the field lines don't begin on that source. Instead, they wrap around that source.

- The magnetic field of a long, straight wire consists of concentric circles wrapping around the wire, and the magnetic field of a long, straight wire falls off with distance as the inverse of the distance from the wire.

- The field is falling off in only 2 dimensions instead of 3, so it falls off more slowly than an inverse-square field from a point-like object.

- Incidentally, the electric field of a long, straight charged wire also falls off inversely with distance, although that field points radially outward from the wire, which is its source, and wraps around the conductor.

- The stronger the current, the stronger the magnetic field; if the current were zero, there'd be no magnetic field.

- The constant μ_0 is called mu naught or mu subzero, and it is essentially a magnetic analog of the coulomb constant k. This magnetic constant has the value $4\pi \times 10^{-7}$ tesla-meters per ampere (T·m/A).

- Because currents respond to magnetic fields and because currents give rise to magnetic fields, if you have 2 wires in each other's vicinity, they will give rise to magnetic forces on each other. One current gives rise to the magnetic field, and the other current gives rise to the response.

- The simplest electrical entity is the point charge. It has a radially outward field and falls off as $1/r^2$. You can take 2 point charges and put them together to create a dipole.

- The simplest magnetic entity, in contrast, is the current loop, which is basically a dipole that makes a dipole field that falls off as $1/r^3$.

- In the electric case, fields are generated directly by electric charges. In the magnetic case, fields are caused by moving charges, and moving charges respond to magnetic fields.

- In magnetism, the moving electric charges are atomic. The current associated with a particular electron is flowing in the opposite direction, and the tiny atomic current loop creates a magnetic dipole.

- The magnetic dipole moment associated with an atomic current loop would cause the loop to experience a torque in the presence of a magnetic field.

- Electrons and other subatomic particles have a property called spin—you can very crudely think of it as a little ball that's spinning.

- Because the electron has charge, it also constitutes a miniature current loop; consequently, it also has a magnetic dipole moment that could either point in the same direction as the one associated with the orbital motion of the electron, or it could point in the opposite direction.

- Basically, atoms become current loops, and that means all matter have some magnetic properties. Iron and steel, for example, may be the most dramatically magnetic, but all matter has magnetic properties.

- There are 3 types of magnetism; the most familiar type is **ferromagnetism**.

- In ferromagnetic materials, there is a very strong interaction among nearby atomic magnetic dipoles that causes them all to align in the same direction. These materials develop relatively large, multiple-atom domains in which all the atomic dipoles are pointing in the same direction.

- In another domain, the atomic dipoles might be pointing in a different direction, so a random piece of ferromagnetic material—of which iron and iron-containing compounds are most common—is not magnetized itself, but it contains these magnetic domains.

- To make a material into a permanent magnet, you can magnetize it by forcing all the domains to line up by, for example, putting it in a strong magnetic field.

- A permanent magnet will attract any ferromagnetic material because it will result in the dipoles in the ferromagnetic material, at least temporarily aligning, and then it also becomes a magnet.

- The south pole is formed, attracted to the north pole of the magnet you begin with, and you get the everyday phenomenon of magnetism—so there's a very strong attraction of ferromagnetic material to magnets.

- Another form of magnetism is called **paramagnetism**, which is a lot less common. In paramagnetism, the individual atomic dipoles don't tend to align very strongly.

- There is a weak interaction among the dipoles, so there is a very weak attraction to ferromagnets. We can measure paramagnetism, but it's not something you're going to see in everyday life.

- Finally, **diamagnetism** occurs when a magnetic field changes near the atomic dipoles, and they actually respond by developing a magnetic dipole moment that causes them to be repealed from magnets.

- Diamagnetism is the opposite of paramagnetism; it's a weak interaction, but it's a repulsive interaction.

- Diamagnetic materials are repelled from magnets. The magnetic moments have to be induced by changing magnetic fields as, for example, a magnet approaches a diamagnetic material.

Important Terms

Biot–Savart law: States that a very short length of current produces a magnetic field that falls off as $1/r^2$: $\Delta B = \dfrac{\mu_0}{4\pi}\dfrac{I\Delta Lr}{r^2}$.

diamagnetism: The opposite of paramagnetism; it's a weak interaction, but it's a repulsive interaction. It occurs when a magnetic field changes near the atomic dipoles, and they respond by developing a magnetic dipole moment that causes them to be repealed from magnets.

Lecture 35: The Origin of Magnetism

ferromagnetism: This is the common, everyday magnetism that is familiar to us. In ferromagnetic materials, there is a very strong interaction among nearby atomic magnetic dipoles that causes them all to align in the same direction.

magnetic monopole: A beginning or end of field lines (e.g., a positive charge is always at the beginning of electric field lines). This concept is important because nobody has ever found a magnetic monopole; therefore, magnetic field lines can never begin or end—they must form loops.

paramagnetism: A type of magnetism that is less common than ferromagnetism in which the individual atomic dipoles don't tend to align very strongly.

Suggested Reading

Rex and Wolfson, *ECP*, chap 18.4–18.6.

Wolfson, *EUP*, chap 26.5–26.8.

Questions to Consider

1. What is the fundamental source of magnetic fields?

2. Will you ever see a magnetic field whose configuration resembles the electric field of an electric point charge? Explain.

3. What makes ferromagnetic materials so much more obviously magnetic than other materials?

Electromagnetic Induction
Lecture 36

T he intimate relationship between electricity and magnetism extends to the fields themselves. Electric charge is only one source of the electric field; another source is changing magnetic field. The relative motion of a magnet and an electric conductor induces current in the conductor. Faraday's law describes this phenomenon of electromagnetic induction, which, at its most fundamental, involves creation of an electric field by a changing magnetic field. Electromagnetic induction is consistent with the principle of conservation of energy.

- Electromagnetic induction gives a direct relationship between the electric and magnetic field.

- A changing magnetic field induces a current in a nearby complete circuit; more specifically, a changing magnetic field creates an emf, a source of energy in an electric circuit that is approximately the same as voltage.

- If you have a conductor and a changing magnetic field arises near that conductor, there will be a voltage that drives a current if it is a complete circuit. More fundamentally, a changing magnetic field creates an electric field.

- There are 2 ways of making electric fields: You can have electric charges that produce electric fields in the space around them, or you can have changing magnetic fields that also produce electric fields.

- An electric field created by a changing magnetic field is not a conservative field—unlike the one produced by electric charges.

- The statement that changing magnetism makes electricity is embodied in one of the fundamental laws of electromagnetism called **Faraday's law**, which says that voltage induced in a circuit is the rate of change of what is called magnetic flux through the circuit.

- **Magnetic flux** is the product of the field with the circuit area. The symbol for magnetic flux is the Greek letter phi with the subscript *B*, standing for magnetic fields: Φ_B.

- The statement of Faraday's law in this context says the magnitude of the induced voltage is the rate of change of the magnetic flux divided by the rate of change of time: $\varepsilon = \dfrac{\Delta\Phi_B}{\Delta t}$.

- Magnetic flux is a surrogate for the number of field lines emerging or passing through a surface bounded by the given circuit.

- If the field is uniform, you can calculate magnetic flux by multiplying area by magnetic field. If it is nonuniform, you have to use some calculus.

- There are 3 ways to change the magnetic flux through a circuit: You could change the field, the area, or the orientation of the area relative to the field.

- Faraday's law addresses the magnitude of the induced magnetic field, but we also need to know the direction.

- **Lenz's law** states that the direction of any induced voltage or current opposes the change causing it, giving rise to the induced set.

- If Lenz's law were not true, energy conservation would not hold.

- It is essential for electromagnetism to be consistent with energy conservation, and mathematically, Lenz's law is expressed by a minus sign in Faraday's law.

- In practice, we calculate the magnitude of the induced voltage, and then we figure out what the direction should be from the fact that the direction of the induced current has to be opposite the induced effect to try to stop the induced effect from occurring.

- Electromagnetic induction is consistent with the law of conservation of energy as we learned when we studied mechanics.

- To prove this statement, we have a simple system in Figure 36.1 consisting of a couple of electrically conducting rails—picture a railroad track, for example—that are spaced a distance apart and are spiked into the ground.

Figure 36.1

Conducting rails, spacing *d*
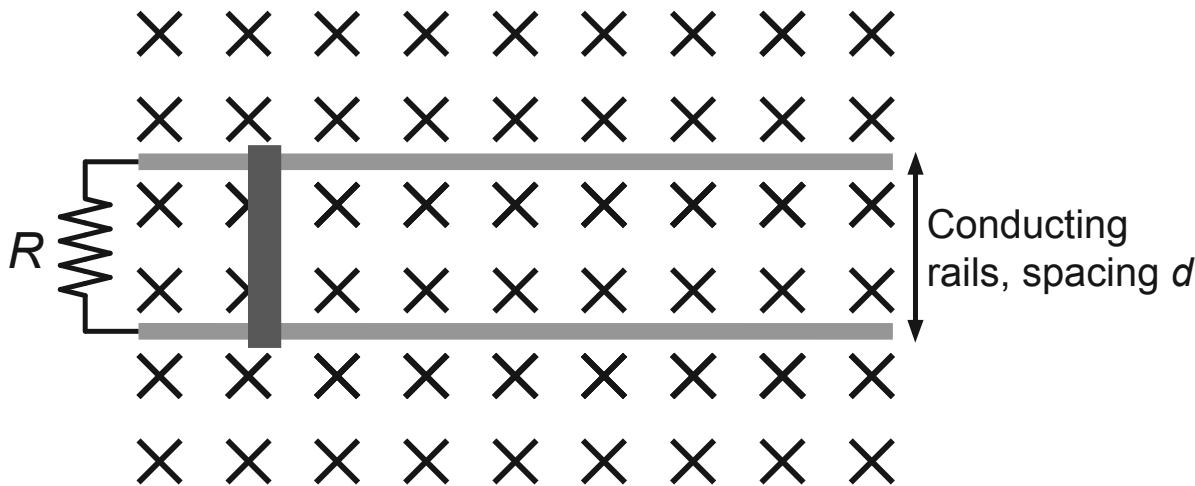
- Across the end of them a resistance *R* is connected; they are electrically insulated from each other except at this resistance. The whole system is sitting in a magnetic field, which is pointing into the page of this guidebook.

- There is a vertical bar that slides along the rails and is greased, so it is frictionless. The bar is also a conductor, and it is in electrical contact with the rails.

• This is a complete electrically conducting loop; we can assume there is not resistance anywhere except in the resistor.

• You pull the bar to the right with a constant speed **v**. There is electromagnetic induction, so you want to know the rate at which you have to do work on that bar and the rate at which electrical energy is dissipated in the resistor.

• These 2 wonderings are exactly the same: This system takes the mechanical work that you do by pulling on the bar, and it converts it ultimately into heat in the resistor.

• The loop area, at least the part of it that is in the magnetic field, is the rail spacing d times some distance x. It is a rectangular area, so $A = dx$.

• Because the magnetic flux is the magnetic field—in this case uniform—times the area, it becomes **B**dx because dx is the loop area.

• Induction is about changing magnetic flux, so we do not actually want the magnetic flux, we want the rate of change of magnetic flux.

• As the bar moves, the area bounded by the conducting circuit is going to increase. Consequently, the flux increases. The magnetic field, in this case, is not changing.

• We have to find the rate of change of the area to find the rate of change of the flux, which is what goes in Faraday's law to give us the voltage that is induced.

• The area is not changing either. What is changing is x, which is causing the change in flux, which is then making the area change. The rate of change of x is the velocity **v**.

- This expression $\dfrac{\Delta \Phi}{\Delta t} = B\dfrac{\Delta A}{\Delta t} = Bd\dfrac{\Delta x}{\Delta t} = Bdv$ becomes v, which is the rate of change of position. This is also the rate of change of magnetic flux.

- Faraday's law states that the magnitude of the voltage is the rate of change of magnetic flux, so the magnitude of the voltage is Bdv:
$$|v| = \frac{\Delta \Phi}{\Delta t} = Bdv.$$

- Ohm's law tells us the current, which is $I = V/R = Bdv/R$. The direction of the current is counterclockwise because Lenz's law states that the direction of the current has to oppose the change that is causing it, which is the increase in magnetic flux.

- A current can try to slow the increase in magnetic flux by making a magnetic field that opposes the magnetic field that is associated with the induced effect.

- The induced current flows in such a direction that its magnetic field in the center of the loop area is going to oppose the magnetic field that points into the page.

- Induced effects do not always oppose magnetic fields; they oppose changes in magnetic flux. In this case, the direction is opposite because the magnetic flux is increasing.

- Because $I = (Bdv)/R$ and $V = Bdv$, the electrical power $P = IV = (Bdv)^2/R$. That is the rate at which electrical energy is dissipated in the resistor.

- We found that the magnetic force on the current-carrying wire is the current times the length of the wire—in this case, d times the magnetic field—so $F = IdB$ to the left.

- The force you need to apply for constant velocity would be a force in the opposite direction to make zero net force, so the velocity would be constant. You need to apply a force of magnitude IdB to the right, but we know $I = Bdv/R$, so you have B^2d^2v/R to the right.

- Force times distance is work, so force times distance per time is rate of doing work, or power: $P = Fv = (Bdv)^2/R$.

- Therefore, the rate at which electrical power is dissipated in the resistor is exactly equal to the rate at which work was done, so electromagnetic induction takes the work and turns it into electrical heating of the resistor.

Important Terms

Faraday's law: A field equation for electromagnetism that describes how electric fields curl around changing magnetic fields (electromagnetic induction).

Lenz's law: States that the direction of any induced voltage or current opposes the change causing it, giving rise to the induced set.

magnetic flux: The net amount of a field that flows through a surface. The concept is directly related to flow for the wind field but can be extended by analogy to electric and magnetic fields.

Suggested Reading

Rex and Wolfson, *ECP*, chap 19.1, 19.2.

Wolfson, *EUP*, chap 27.1–27.3.

Questions to Consider

1. A magnetic field points into the page of this guidebook, and it's decreasing in strength. If a wire loop lies in the plane of the page, what will be the direction of the induced current in the loop?

2. Solar activity can cause fluctuations in Earth's magnetic field, which can wreak havoc with power lines. How is this possible?

3. You push a bar magnet toward a conducting loop; an induced current flows, heating the loop. What's the ultimate source of the energy that heats the loop?

Applications of Electromagnetic Induction
Lecture 37

I nduction is the basis of many important technologies, including the electric generators that supply nearly all the world's electrical energy; transformers that let us adapt electrical energy to specific uses; security devices such as metal detectors; microphones that convert sound to electrical signals; electric guitars; induction stovetops; and many more. Induction is used to read the information stored on magnetic media, including the use of machines we use daily that swipe credit cards.

- Electromagnetic induction is the source of nearly all of the world's electrical energy. We generate electricity with electric **generators**, which involve changing magnetic flux by changing the orientation of a coiled wire in a magnetic field.

- A generator and a motor are essentially the same—but operated in reverse. Put electricity into a motor, and you get mechanical energy out. Put mechanical energy into a spinning coil of wire in a magnetic field, and electromagnetic induction gives you electricity out.

- About 40% of the world's primary energy consumption goes into making electric power, although the process is not very efficient; only about 12% of the energy the world actually uses is in the form of electricity, and almost all of it is generated in electric generators.

- In principle, you would like generators to be 100% efficient at generating energy; they are not 100% efficient, but they are pretty efficient. The inefficiencies in power plants occur in the process of converting thermal energy into electricity.

- To design an electric generator, we start with a simple loop of wire in a magnetic field B. The wire loop is initially oriented so it is perpendicular to and in the same direction as the magnetic field. The angle between them (θ) is zero.

- As time goes on, the loop is going to rotate. The flux through one turn of that loop is going to be the magnetic field times the area times the cosine of the angle between them: $\Phi = BA\cos\theta$.

- As the angle increases, the flux decreases for a while because the number of field lines going through the area is decreasing.

- When the loop is perpendicular to the field, the flux is zero, and then it gets bigger again as the loop continues to go around.

- In rotational motion, theta increases linearly with time, and the rate of increase is omega (ω), the angular velocity: $\theta = \omega t$.

- The rate of change of cos (ωt) is $-\omega\sin(\omega t)$.

- As an example, we can plug in the value of 2 for omega and get the cos($2t$), which can be graphed in conjunction with 2sin($2t$) to show their maximum and minimum points. (The negative sign is not necessary when analyzing the graphs of these functions.)

- Where the cosine curve is changing most steeply is where the sine curve has its biggest values.

- The cosine curve is flat at its maxima and minima, and it is not changing instantaneously. It is at these maxima and minima that the rate of change goes through zero.

- The cosine function turns into a sine function when you figure out its rate of change, and it is multiplied by the angular frequency because the more rapidly it is changing, the bigger the rate of change.

- There is a 500-turn loop that has an area of 0.018 m^2. It is spinning at 60 revolutions per second (rev/s), which will make it generate 60-Hz alternating current. It is also in a 50-mT magnetic field.

- What is this generator's peak output voltage? The induced voltage is always the rate of change of the magnetic flux, so $|V| = \dfrac{\Delta\Phi}{\Delta t} = NAB\omega\sin\omega t$. The minus sign is not included because we need the absolute value.

- The angular velocity is 2π times the frequency because there are 2π radians in a full circle. The frequency is 60 rev/s, so the angular frequency is 377 rad/s: $\omega = 2\pi f = (2\pi)(60 \text{ rev/s}) = 377$ rad/s.

- Spinning coils in magnetic fields naturally make sinusoidally varying alternating currents, and the maximum value of the sine function is 1. The value of the generator's peak output is $V_{peak} = NAB\omega = (500)(0.018\text{m}^2)(0.05 \text{ T})(377 \text{ rad/s}) = 170$ V.

- We use devices called **transformers** to get the high voltages that we use to transmit electric power. We use transformers widely throughout the power system to alter voltage levels with alternating current.

- Because alternating current makes changing magnetic fields and changing magnetic fields make electromagnetic induction, we can easily change voltage levels by using transformers.

- In transformers, a primary coil produces changing magnetic flux, which induces voltage in a secondary coil. Then, an iron core concentrates the magnetic flux.
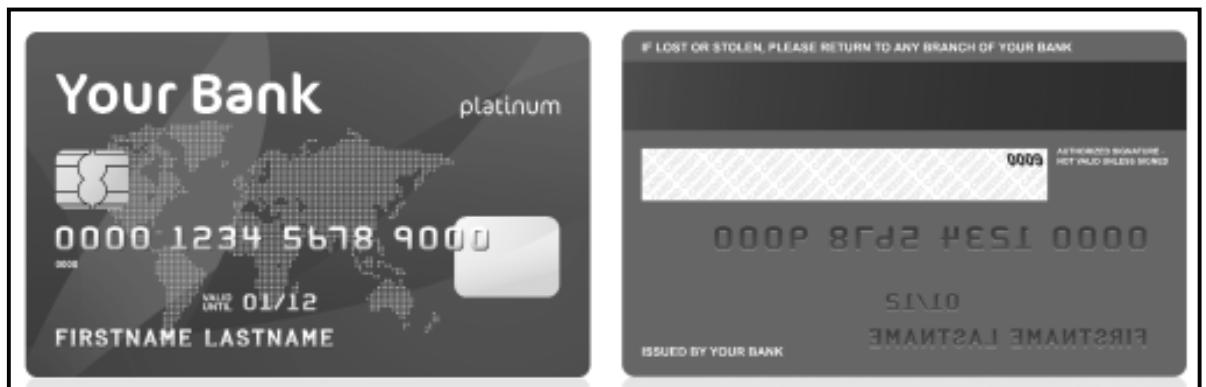
When you swipe your credit card, the magnetic strip passes a coil that is wound around an iron core, and a current is induced that takes information off the card.

- In principal, the output voltage should be equal to the input voltage as the ratio of the turns in the 2 coils. However, not quite all of the magnetic flux is transferred—transformers are not perfect, but we can make them fairly efficient.

- Electromagnetic induction is not just about making currents flow in particular circuits. Currents can also flow in conductive material that is subject to changing magnetic fields.

- **Eddy currents**, currents in conductive material caused by changing magnetic fields, dissipate rotational kinetic energy.

- In eddy-current breaking, there are spinning conducting materials—like the break hub on a high-speed train, a Japanese bullet train.

- In this example, an electromagnet turned on right near the break hub induces eddy currents, which sap the rotational kinetic energy, and the train comes to a stop.

- Japanese bullet trains used that kinetic energy to turn electric generators and put the energy back in a battery instead of dissipating it as heat.

- Magnetic strips are used on the back of credit cards; those strips are swiped past a little coil wound around an iron core, and currents are induced that take off the information that is contained in that card.

- The speed with which you swipe a credit card cannot matter, and yet because induction depends on the rate of change of magnetic flux, the faster you swipe the card, the bigger the induced effects.

- The information on the credit card has to get stored in varying patterns, which will appear no matter how big the voltage gets or how close the patterns get together.

- Every time you swipe a card, think about Faraday's law and electromagnetic induction because that is what is doing the reading.

- There are countless devices in our everyday lives that use electromagnetic technology: electric power generators, transformers, transducers (microphones), eddy currents flowing in conductive materials, security devices (metal detectors), and information technologies—especially the swiping of credit cards.

Important Terms

eddy current: A current in conductive material caused by changing magnetic fields that dissipates rotational kinetic energy.

generator: A device that uses electromagnetic induction to convert mechanical energy to electrical energy. Typically, a generator involves a coil of wire rotating in a magnetic field.

transformer: A device that uses electromagnetic induction to transform high-voltage/low-current electricity to low-voltage/high-current, and vice versa.

Suggested Reading

Rex and Wolfson, *ECP*, chap 19.3.

Wolfson, *EUP*, chap 27.3.

Questions to Consider

1. Why do electric generators naturally produce alternating current that varies sinusoidally with time?

2. How does the reader of a credit card extract information from the card?

3. Why do transformers only work with alternating current?

Magnetic Energy
Lecture 38

E|lectric fields and currents produced by electromagnetic induction act to oppose the changing magnetic fields that give rise to them. Induction therefore makes it difficult to change the current in a circuit, as the changing current creates a changing magnetic field that, in turn, induces an electric field that opposes the current change. This effect is called self-inductance, which is useful technologically, but it can also be hazardous. Together, electricity and magnetism have many complementary aspects, including the storage of energy in either type of field.

- Suppose there is no current in a circuit, but you want to start a current flowing around a wire loop. As soon as you start a current flowing, there is associated with that current a magnetic field because moving electric charge is the source of magnetic fields.

- The magnetic field lines wrap around the wire that is carrying the current, and in particular, the magnetic field lines penetrate through the area bounded by the circuit.

- There is a magnetic flux through this circuit that is caused by the circuit's own magnetic field, caused by the current that is flowing in that circuit.

- Changing magnetic flux creates electric fields and that those electric fields oppose the change that is giving rise to them.

- If you are trying to build up the current in the loop starting at zero, as soon as you try to change the current, you create a changing magnetic flux in the circuit itself, which is going to oppose the change that gave rise to it—mainly, the buildup of current.

- The net result is it becomes difficult to build up currents and circuits. This is a property of all circuits that is called **self-inductance**.

- The property of self-inductance is most important in circuits that enclose a lot of their own magnetic flux. In principal, this property exists in every circuit and is especially important when you are working with high-frequency circuits like those in computers and televisions.

- If a circuit encloses a certain amount of its own magnetic flux, self-inductance is the ratio of the magnetic flux (Φ_B) to the current—the more current, the more magnetic field, the more flux: $L = \Phi_B/I$.

- This ratio does not depend on either the flux or the current, but it does depend on the geometrical configuration of the circuit.

- The unit of flux is the tesla-meter squared per ampere (T·m/A)—flux per ampere per current—which is given the name henry (H) after the American scientist Joseph Henry.

- Inverting the definition, the flux is the self-inductance times the electric current that is flowing in that circuit: $\Phi_B = LI$.

- If you take the rate of change of both sides of that equation, the rate of change of magnetic flux is the inductance times the rate of change of the current in the circuit: $\Delta\Phi_B/\Delta t = L(\Delta I/\Delta t)$.

- The self-inductance L is a property of the circuit that does not change. The only thing that can change is a change in current on the right-hand side of the equation.

- A device that is specifically designed to have a particular value of inductance—typically a coil of wire—is called an **inductor**.

- Based on Faraday's law, from the principle of electromagnetic induction, the induced voltage is minus the rate of change of the magnetic flux: $\varepsilon_L = -L(\Delta I/\Delta t)$.

- Combining the principle of electromagnetic induction and the definition of self-inductance, you find that there is a voltage across the inductor, and it is given by Faraday's law minus the rate of change of the magnetic flux.

- You also find that the rate of change of magnetic flux was the inductance times the rate of change of current.

MRI scanners are large solenoids, which are used to create straight-line motion in electrical-mechanical systems.

- There is a voltage across an inductor, which is related not to the current through it as in a resistor, but to the rate of change of the current because of the component $-L(\Delta I/\Delta t)$.

- In this equation, $\varepsilon_L = -L(\Delta I/\Delta t)$, the inductor voltage is minus the inductance times the rate of change of current. On the left-hand side is the symbol for an inductor, which appropriately looks like a coil of wire.

- The voltage travels across the inductor, and the current travels through the inductor—in this case, flowing downward.

- The voltage is defined to be positive if it increases as you go in the direction of the current and negative if it decreases.

- There is a voltage across the inductor, the inductance, and the rate of change of current, which are the elements found in the anatomy of the equation that describes inductor voltage: $|V| = L(\Delta I/\Delta t)$

- The consequence of all of this is that the current through an inductor cannot change instantaneously.

Lecture 38: Magnetic Energy

- It is difficult to build up the current, and it is also difficult to stop the current because it will not stop flowing instantaneously. If it did, there would be an infinite $\Delta I/\Delta t$ and an infinite voltage, which is impossible.

- The convergence of electricity and magnetism into the unified field of electromagnetism is going to come to complete fruition in Lecture 40, but we are approaching it in this lecture. There is an important complementarity between electricity and magnetism.

- In the electricity case, we have parallel-plate capacitors that create uniform electric fields. There is stored energy $U = \frac{1}{2}CD^2$ inside a capacitor, and it is stored in the electric field of the capacitor.

- The voltage across those capacitors cannot change instantaneously because it takes time to move charge on and off the plates of the capacitor.

- More profoundly, the energy stored in a capacitor, $\frac{1}{2}CD^2$ stored in a parallel-plate capacitor, is a simple configuration whose capacitance we could calculate.

- The energy density stored in all electric fields has the form one-half times a constant involving the electric constant times the strength of the electric field squared: $u_{\mathrm{E}} = \frac{1}{2}\in_0 E^2$.

- In magnetism, we have long solenoid coils. A solenoid looks nothing like a parallel-plate capacitor, and it does not work on anything like the same principles, but it is complementarily analogous—it is the magnetic analog.

- Considering the effect of self-inductance in a solenoid coil shows that magnetic field within the coil is a repository of stored energy—magnetic energy, analogous to the electric energy stored in electric fields.

- A solenoid creates a uniform magnetic field and has stored energy $(1/2)LI^2$, with the inductance L playing the role of capacitate C. The current I plays the role of voltage V, and the current cannot change instantaneously.

- If you try to make the current change instantaneously, you can have disastrous results because it leads to high induced voltages in circuits that are subject to rapidly changing currents.

- Finally, we reach the profound conclusion that there is energy density in magnetic fields. Energy density is one-half times a constant—the magnetic constant, or one divided by it—times the strength of the electric field squared: $u_B = (1/2)(B^2/u_0)$.

- These mundane, practical devices—capacitors and inductors—have helped us to realize the profound complementarity that exists between electricity and magnetism.

- This near-equality or analogy between electricity and magnetism pervades throughout the entire universe and gives electric and magnetic fields one of the most important quantities in physics: energy.

Important Terms

inductor: A device that is specifically designed to have a particular value of inductance. Typically, a coil of wire is used as the inductor.

self-inductance: The property of a circuit that allows the circuit to induce a current in itself.

Suggested Reading

Rex and Wolfson, *ECP*, chap 19.4.

Wolfson, *EUP*, chap 27.4–27.6.

Lecture 38: Magnetic Energy

Questions to Consider

1. How is a solenoid's magnetic field similar to a parallel-plate capacitor's electric field?

2. If you connect a resistor and inductor in a series across a battery, current gradually builds up in the inductor—and with it, magnetic energy. Within the circuit, what's the original source of that energy?

3. How does the energy density in a magnetic dipole's magnetic field fall off with distance from the dipole?

AC/DC
Lecture 39

D irect current (DC) is electric current that flows in one direction; alternating current (AC) flows back and forth. AC is described by its amplitude—the peak level of voltage or current—and frequency, or the number of back-and-forth cycles per second. Capacitors and inductors respond to AC by alternately storing and releasing energy, rather than dissipating it into heat as do resistors. Combining a capacitor and inductor in a circuit gives an oscillator. Furthermore, capacitors and inductors extend the complementarity between electricity and magnetism.

- **Direct current**, or DC, is the current produced, for example, by batteries. Power supplies convert **alternating current**, or AC, coming out of the wall into direct current.

- Almost all of our power systems use alternating current—partly because we can easily transform AC using the law of induction and transformers, devices that allow the long-distance transmission of power.

- An alternating voltage (AC voltage) is associated with an alternating current, but we will just use the term AC, even though it stands for alternating current, to describe either voltage or current.

- In North America, the frequency of that alternating voltage and current is 60 Hz, 60 cycles per second, which converts to about 377 radians per second.

- The graph for AC voltage is a sinusoidal variation: $I = \dfrac{V}{R} = \dfrac{V_{\mathrm{p}} \sin \omega t}{R} = \dfrac{V_{\mathrm{p}}}{R} \sin \omega t,$ where the peak voltage is V_{p} and the angular frequency is ω.

- The angular frequency ω is $2\pi f$. The root-mean-square voltage V_{rms}, an average value, is $I_p / \sqrt{2}$.

- The relative timing of 2 AC quantities is called phase, which is characteristic of 2 different AC signals and refers to when they peak in relation to each other.

- The voltage in AC circuits is $V_p \sin(\omega t + \varphi_I)$, with φ representing the phase constant, but its phase is zero because it starts rising right at time equals zero, so the voltage becomes $V_p \sin(\omega t)$.

- The current in AC circuits is $I_p \sin(\omega t + \varphi_I)$.

- If you were to calculate the average power in a particular situation, sometimes the voltage would be positive and the current negative, and the power would be negative. During that time, the circuit would actually be giving back power to its source.

- At other times, power would be positive, and if you work that out over the whole cycle, you would find that there is no net power consumed over an entire cycle.

- To send power, you have to have voltage and current in phase. As soon as they slip out of phase, you are sending less power for a given current than you could otherwise by having them in phase. This notion plays an enormous role in some major power failures.

- Voltage and current get out of phase as a result of the presence of either capacitance or inductance in the circuit.

- Resistors are the simplest kinds of components. The relationship between current and voltage in a resistor, as given by Ohm's law, says the bigger the voltage for a given resistance, the bigger the current will be, and the bigger the resistance for a given voltage, the smaller the current will be.

- The fundamental defining relationship for a capacitor is its capacitance, or the charge divided by voltage—rewritten as $Q = CD$.

- There is not a direct relationship between current and voltage in a capacitor; there is a relationship between charge and voltage in a capacitor.

- The rate of change on the capacitor plates, rate of change on the charge, is the rate at which charge is flowing through those wires, and that is simply called current.

- The equation that describes the capacitor's current-voltage relationship involves time and says that the current is proportional to the rate of change of voltage—rather than a direct relationship between current and voltage.

- It takes current to move a charge onto a capacitor, so current has to start flowing to a capacitor before the capacitor can build up voltage because the voltage across a capacitor is proportional to the charge.

- The higher the frequency, the more rapid the charge movement is to get charge on and off the plates and, therefore, the higher the current becomes.

- The peak current is the peak voltage divided by the quantity one divided by the frequency times the capacitance.

- The capacitor acts like a resistor, with the resistance being one divided by the quantity frequency times capacitance, and that quantity is called the **capacitive reactance**: $X_C = 1/\omega C$.

- This quantity, $1/\omega C$, is not the same as resistance because there is also a phase lag introduced by a capacitor in which current leads voltage.

- In inductors, the voltage is the inductance times the rate of change of current. Voltage is related to rate of change of current, and the capacitor current is rate of change of voltage.

- The induced voltage arises as soon as the current starts to change—it was initially zero—and if you try to build it up, you get voltage immediately.

- The voltage leads the current in an inductor, and it does so by $\pi/2$ radians or 90°, both of which equate to a quarter of a cycle.

- An inductor is the opposite of a capacitor: The higher the frequency, the more rapidly the current is changing, and the bigger the voltage that tries to keep the current from changing—which leads to a smaller current.

- An inductor sort of acts like a frequency-dependent resistor with its resistance equaling the frequency times the inductance.

- However, the whole picture of what the inductor does is to limit the flow of current in a f requency-dependent way, but also to introduce a phase lag.

- A resistor has a direct proportionality between voltage and current; a capacitor has a direct proportionality between charge and voltage; and an inductor has a direct proportionality between magnetic flux and current.

- In an *LC* circuit (an inductor-capacitor circuit), the current peak and the reactance will cancel when the inductive reactance and the capacitive reactance are equal—that is, when ωL, the inductive reactance, equals $1/\omega C$, the capacitive reactance.

- Algebraically, $\omega = 1/\sqrt{LC}$ which is the frequency at which the current will be a maximum.

- At that frequency, called the natural frequency, the inductor and capacitor voltages have cancelled out, and as far as the generator is concerned, it is just a circuit with a pure resistance.

- This phenomenon, whereby the current is a maximum with a particular combination of capacitance and inductance, should remind you of something we talked about in Lecture 17—namely, resonance and mechanical systems—whereby if a system were driven at just the right frequency, it would develop large-scale oscillations, which could have disastrous consequences.

- The same thing happens in electrical circuits, but the difference is we can make a much broader range of oscillations electrically than we can mechanically.

- The inductance and capacitance set a frequency at which the circuit has the greatest response. The sharpness of that resonance depends on how much resistance there is in the circuit.

Thomas Edison introduced the first electric power grid in the 1890s in New York City.

Library of Congress, Prints and Photographs Division, LC-USZ62-98066.

- Capacitors and inductors behave complimentarily with frequencies of oscillation, which are set by the conductor and capacitor in the electric case and by the spring constant and the mass in the mechanical case.

Important Terms

alternating current (AC): Electrons in a circuit oscillate back and forth instead of flowing. (Compared with DC, or direct current.)

capacitive reactance: When a capacitor acts like a resistor, the resistance is one divided by the quantity frequency times capacitance: $X_C = 1/\omega C$.

direct current (DC): Electrons in a circuit flow in only one direction. (Compared with AC, or alternating current.) DC would result from a circuit with a battery; AC would result in household circuits.

Suggested Reading

Rex and Wolfson, *ECP*, chap 19.5–19.6.w

Wolfson, *EUP*, chap 28.

Questions to Consider

1. In what ways are inductors and capacitors complementary devices? What deeper complementarity does this reflect?

2. A capacitor is connected across an AC generator. Over one AC cycle, how much net energy does the capacitor take from the generator? Contrast with the case of a resistor.

3. *LC* circuits used in radio and TV tuning need to be highly selective, meaning they can readily distinguish nearby channels. Should these circuits be designed to have high or low resistance? Explain.

Electromagnetic Waves
Lecture 40

C hanging magnetic fields give rise to electric fields. In the 1860s, Maxwell completed the set of 4 equations describing electromagnetic fields by adding a term that makes changing electric field a source of magnetic field. That led to the possibility of electromagnetic waves. Maxwell calculated the speed of these waves and found it was the known speed of light, concluding that light is an electromagnetic wave. There is an entire spectrum of electromagnetic waves that are distinguished by frequency and wavelength. Electromagnetic waves originate in accelerated electric charge and carry energy.

- Implicit in everything we have said about electromagnetic fields, there are 4 fundamental and different statements about how electromagnetic fields behave.

  ○ First, electric fields arise from charges from Coulomb's law. Its geometrically very different but equivalent law, Gauss's law, tells us how electric fields arise from charges and, in particular, that electric field lines begin and end on charges.

  ○ Magnetic fields arise from moving electric charges because of the Biot–Savart law and Ampère's law.

  ○ Magnetic field lines do not begin or end, and there is no magnetic analog of electric charge. There is no such thing as a magnetic monopole; in electricity, there are monopoles, electric charges that give rise to electric fields.

  ○ Finally, changing magnetic fields give rise to electric fields because of Faraday's law of electromagnetic induction.

- In the 1860s, James Clerk Maxwell, a Scottish physicist and mathematician, had a brilliant insight about these 4 equations.

- Maxwell noted that Faraday had said that changing magnetic fields give rise to electric fields, which is the idea of electromagnetic induction.

- Maxwell wondered whether changing electric fields could give rise to magnetic fields.

- Specifically, Maxwell added a changing electric flux to Ampère's law—the same way there is a changing magnetic flux in Faraday's law.

- This addition then became a new source of magnetic field—just like changing magnetic flux in Faraday's law became a new source of electric field.

- Ampère's law, which originally talked about the magnetic field's curling property being related to electric currents, obtains another term to its equation. In other words, magnetic field arises from electric current, and then it also arises from changing electric flux.

- In the SI system of units, the changing electric flux term has in front of it a constant, which incorporates both the electric and magnetic constants together in one term for the first time.

- By adding this term, the problem of conservation of electric charge with unsteady currents was solved.

- Maxwell's equations in vacuum describe the electric and magnetic fields that exist in vacuum, which is not devoid of all things, as it might seem to be.

- Maxwell's brilliant insight led to the symmetry between Faraday's law and Ampère's law. In the absence of electric charge, magnetism and electricity stand on completely equal ground; their equations are completely symmetric in vacuum.

- Maxwell realized that there is a kind of self-replicating electromagnetic structure propagating through empty space as each changing field gives rise to the other.

- It is not essential that a changing magnetic field make a changing electric field, but unless the changing magnetic field is changing uniformly—and that is impossible because it would just have to keep growing forever—then the electric field gets induced and is going to be changing, and it will give rise to magnetic fields that will also be changing.

Cell phones, and other wireless technological devices, operate by the use of electromagnetic waves.

- Therefore, change in an electric field gives rise to a magnetic field, which itself is likely to be changing. The change in the magnetic field gives rise to an electric field, and that process repeats itself so that electric and magnetic fields are continually regenerating each other. This structure is called an **electromagnetic wave**.

- The electric and magnetic fields are perpendicular to each other, and they are also perpendicular to the direction the wave is going. That means that these electromagnetic waves in vacuum are transverse waves, just like the ones we talked about in Lecture 18.

- The electric and magnetic fields are also in phase, which means that their peaks occur at the same point in time and that they go through zero at the same point in time.

- The wave speed is determined by Faraday's and Ampère's laws, which are making this continually self-replicating electromagnetic structure.

- The wave speed, the speed of an electromagnetic wave in a vacuum, is $1/\sqrt{\epsilon_0 \, \mu_0}$, which is also equivalent to 3×10^8 meters per second, which is the known speed of light.

- Maxwell discovered that light is an electromagnetic wave. Because of this realization, he had suddenly brought the whole science of optics under the umbrella of electromagnetism.

- One of the greatest intellectual achievements in the history of science was the recognition that light is an electromagnetic wave because light travels at the speed 3×10^8 meters per second, and indeed, it can consist of electric and magnetic fields.

- Electromagnetic waves are not just light, but they are light. They are radio waves, for example, and all kinds of other waves.

- In fact, waves can have any amplitude, but there is a relationship between the strengths of the electric and magnetic fields: The strength of the electric field is the speed of light times the strength of the magnetic field.

- Any frequency is allowed, but as for any wave, the product of frequency and wavelength is the wave speed c: $f\lambda = c$.

- There is a spectrum of all possible wavelengths and frequencies for electromagnetic waves, of which visible light is only a small part.

- At relatively low frequencies and long wavelengths are radio waves. Infrared, then visible light, follow with higher frequencies and shorter wavelengths. At higher frequencies, there are ultraviolet rays, X-rays, and gamma rays.

- There is no firm boundary for the electromagnetic spectrum; these examples are all assemblages of electric and magnetic fields propagating through space at the speed of light.

- Electromagnetic waves are produced ultimately by accelerated charge, which is formed by shaking an electric charge.

- Radio transmitters, for example, have capacitors and inductors connected into a circuit that cause electric currents to oscillate back and forth at a frequency determined by the inductance and capacitance that sends a current up into an antenna. The oscillating, accelerated charges, as they move back and forth, radiate electromagnetic waves.

- Characteristically, the wavelength of an electromagnetic wave is related to the size of the system that is generating it. The size of the system that responds best to electromagnetic waves is roughly the scale of the wavelength.

- In a vacuum, electric and magnetic fields play exactly the same role: $u_E = u_B = [1/(2\mu_0)]B^2$. The complementarity between electricity and magnetism is now complete.

- An electromagnetic wave in vacuum is moving at speed c. If you multiply E and B together—the strength of the electric and magnetic fields, the peak electric and magnetic fields in that wave—and divide by the magnetic constant, you get a quantity called S.

- In vacuum, the quantity S is the average rate of energy flow per unit area and is measured in watts per square meter. It is the rate at which an electromagnetic wave carries energy: $S = EB/\mu_0$.

- Finally, where there is energy there is momentum. The momentum in electromagnetic waves is transferred to whatever the waves hit.

- If the waves absorb the energy, they obtain momentum. If the energy is reflected off the waves, they obtain twice as much momentum, and that phenomenon is called **radiation pressure**.

- Radiation pressure is expressed as a force per unit area and is the intensity *S* divided by the speed of light. Its units are newtons per square meter, and it is relatively weak in most situations.

Important Terms

electromagnetic wave: A structure consisting of electric and magnetic fields in which each kind of field generates the other to keep the structure propagating through empty space at the speed of light, *c*. Electromagnetic waves include radio and TV signals, infrared radiation, visible light, ultraviolet light, X-rays, and gamma rays.

radiation pressure: A phenomenon in which the energy from an electromagnetic wave is reflected off the wave when it comes into contact with an object, and the wave obtains twice as much momentum as a result. This phenomenon is expressed as a force per unit area and is the intensity *S* divided by the speed of light *c*.

Suggested Reading

Rex and Wolfson, *ECP*, chap 20.1–20.2.

Wolfson, *EUP*, chap 29.

Questions to Consider

1. How did Maxwell's modification of Ampère's law restore symmetry to the equations of electromagnetism?

2. Radio waves and X-rays are both electromagnetic waves, with X-rays having much higher frequency. Do X-rays travel faster than radio waves?

3. How does Maxwell's work represent the sort of unification in physics toward which scientists continue to strive?

Reflection and Refraction

Lecture 41

Although light is a wave, we can treat it as if it were traveling in straight lines called rays, which change direction abruptly at boundaries between materials. When light rays reflect from a surface, the incoming and outgoing rays make the same angle with the surface. When light rays refract at the interface between 2 media, they bend toward or away from the perpendicular to the interface, depending on the index of refraction of the 2 media. Snell's law describes refraction mathematically and leads to a phenomenon known as total internal reflection.

- In principle, everything that is reviewed in the next 3 lectures about **optics** can be derived rigorously from Maxwell's equations. Most of the time, however, you can ignore the fact that light is an electromagnetic wave—or that it is a wave at all.

- When light is interacting with objects that are large compared with another wavelength of light, we can ignore the fact that light consists of waves. Instead, we can consider it to consist of straight lines of light that only refract (bend) or reflect when they encounter some other medium. Then, they travel in straight lines again, called rays, within those new media.

- Ray optics is also called geometrical optics because of these straight-line rays that exist.

- In an electromagnetic wave with an electric field, the electric charge of the electron responds to the electric field. The electron oscillates at the frequency of the incoming wave and that oscillating or accelerated electric charge is the source of electromagnetic radiation. The free electrons then reradiate that wave through the process of **reflection**.

234

- An incident ray traveling into a system makes some angle θ_1 with the normal that is perpendicular to the system it is hitting—to the medium off which it is going to reflect.

- The **law of reflection** states that the reflected ray goes out at the same angle relative to the perpendicular at which it came into the system.

- In other words, the **angle of incidence**, the perpendicular angle at which the ray entered, and the angle of reflection are equal.

- In a flat mirror, light from an object shines toward the mirror and bounces off of it, obeying the law of reflection. The light then continues to diverge as it did coming from the object.

- From the viewer's standpoint, it appears that the light is diverging from a point behind the mirror, as far behind the mirror as the actual object is in front. In fact, there is nothing behind the mirror, but it looks like the image is behind the mirror.

- A **virtual image** is an image that is not, in some sense, really there because light is not actually coming from the place that the image is, and it is going to be the same size as the object.

- The image in the flat mirror is going to be the same distance behind the mirror as the object is in front, and it is actually reversed front to back—not left to right—because the part of the actual object facing the mirror corresponds to the part of the image facing the mirror.

- A corner reflector, also called a corner cube, consists of 2 mirrors at right angles in 3 dimensions, and it consists of a third mirror at right angles to those 2.

- In a corner reflector, when light comes in and reflects off one mirror, hits the other mirror, and reflects back, it comes back exactly parallel to how it came in—regardless of what that incoming angle was.

- **Refraction** is the bending of light as it passes from one transparent medium into another.

- Whenever light hits an interface between 2 media, it does not only refract, it also reflects. This statement is true for any wave when it goes into a different medium.

- Different media have different speeds—for light or any other wave. For example, light moves more slowly in glass than it does in vacuum.

- The index of refraction of a material is the ratio of the speed of light to the speed in that material: $n = c/v$.

- In Figure 41.1, medium 1 is in the upper region and medium 2 below, and there is an interface between 2 different media. Wave crests are moving in medium 1 with some speed v_1 such that $n_1 = c/v_1$, which is the definition of the index of refraction.

Figure 41.1
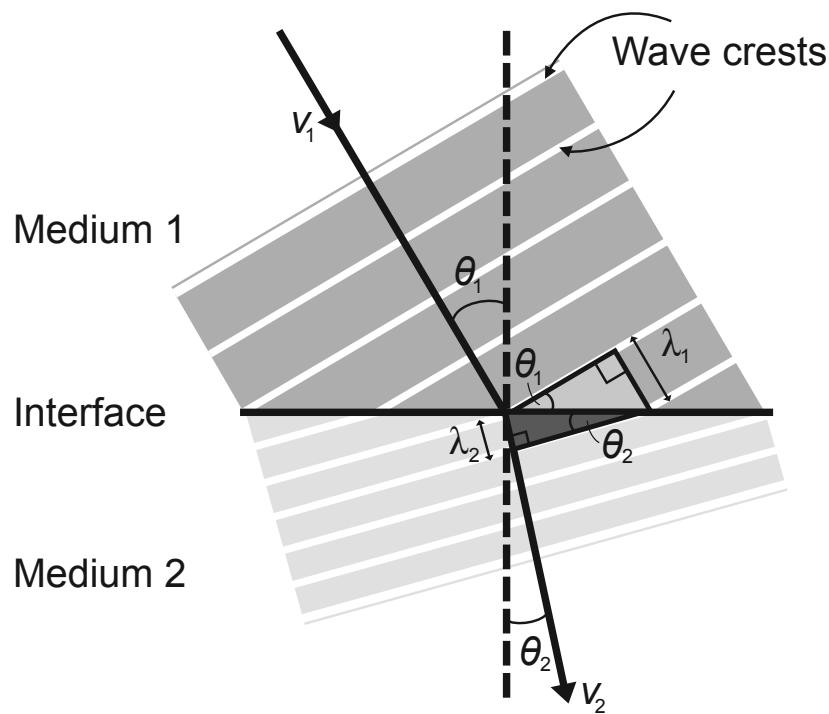
- The wave crests move more slowly in medium 2, so they are slowing down. In order to keep the frequency of the waves the same—the same number of wave crests have to cross that interface coming in as going out—the waves have to slow down because the speed is less, and consequently, the waves bend.

- The speed of a wave is always the frequency times the wavelength: $f\lambda = c$, and the frequency can never change.

- That statement implies that the wavelength in medium 2 is shorter than the wavelength in medium 1 because the speed is $f\lambda$, and f is the same in both cases. The speed is lower in medium 2, so λ has to be lower in medium 2 by that same factor.

- In the figure, there are triangles with angle θ_1 and θ_2. These triangles are right triangles because they both have one side that is perpendicular to the wave crests in each of the 2 different media.

- In addition, these triangles share a common hypotenuse. Remember that the sine of an angle is the opposite side over the hypotenuse.

- The hypotenuse can be written as $\lambda_1/\sin\theta_1$ or $\lambda_2/\sin\theta_2$, and those 2 are equal to each other.

- Substituting the equation for the refractive index into the equation for the wave speed, we get $\lambda = c/nf$.

- This tells us immediately that the wavelength is proportional to the wave speed. The wavelength is smaller in medium 2, where the wave speed is lower.

- Substituting this value, c/nf, into the equation $\lambda_1/\sin\theta_1 = \lambda_2/\sin\theta_2$, we are left with Snell's law after canceling the c's and f's, which are common to both waves: $n_1\sin\theta_1 = n_2\sin\theta_2$.

- When there is almost a perfect reflection, the phenomenon is called **total internal reflection**. It occurs when light is trying to go from a medium that has a higher index of refraction to one with a lower index of refraction.

- If the angle is above some critical angle, the angle of refraction in the lower-index medium would try to be greater than 90°, and that cannot happen, so it reflects totally instead of refracting.

- For glass, the critical angle for total internal reflection is about 42°, and you can calculate the critical angle for any other substance you would like.

- In fact, when we want really good reflection, we will often use total internal reflection instead of simply a shiny coating, which is a property that metals tend to have.

Important Terms

angle of incidence: The perpendicular angle at which a ray enters a system.

law of reflection: States that a reflected ray goes out at the same angle relative to the perpendicular at which it came into a system.

optics: The study of light and how light travels through and between materials. (Geometric optics thinks of light as rays; physical optics tends to think of light as waves—both can be important.)

reflection: The phenomenon whereby a wave strikes a material and rebounds at the same angle with which it struck the material.

refraction: The phenomenon of waves changing direction of propagation when going from one medium to another.

total internal reflection: Complete reflection that occurs as light attempts to go from a more dense to a less dense medium, as from water to air.

virtual image: An image that is not, in some sense, really there because light is not actually coming from the place that the image is.

Suggested Reading

Rex and Wolfson, *ECP*, chap 21.1–21.3.

Wolfson, *EUP*, chap 30.

Questions to Consider

1. What is ray approximation, and under what conditions is it valid?

2. What is the physical reason for refraction at an interface of 2 different optical media?

3. Under what conditions does total internal reflection occur?

Imaging
Lecture 42

C urve a mirror or a piece of glass, and you've got an optical device that bends parallel light rays to a focal point, allowing formation of images. Images can be enlarged or reduced, depending on where the object being imaged is in relation to the focal point. They can be virtual images, meaning light only appears to come from the image, or real images, from which light rays actually come. Image formation is analyzed with ray diagrams or by using mathematical equations.

- A parabolic mirror focuses rays of light that come in parallel to the axis of the parabola and reflect them in such a way that they end up passing through a single point called the focal point.

- If a source of light is placed at the focal point, then the light going out in all directions will bounce off the mirror and form parallel rays, which is how search light beams are made.

- Say there is an object in front of a curved mirror. There are 2 rays: One ray is going in parallel to the mirror axis and goes right through the focus, while the other ray goes through the focus before it hits the mirror and goes out parallel to the axis.

- Where those 2 rays meet determines the image that is produced. If the image is small, inverted, and beyond the focal point, it is called a **real image** because the eye is seeing light that is actually coming from the image.

- **Concave mirrors** form an inverted real image, and the image is in front of the mirror. These types of curved mirrors are devices that reflect light.

- **Lenses** are transparent materials that have curved surfaces, so they control refraction in a way that brings light rays to a focus.

- A lens, unlike a mirror, has 2 focal points. For simple kinds of lenses, the 2 focal distances are the same.

- A **convex lens** is a lens that bends outward. It takes parallel rays and bends them to a focal point.

- Although refraction occurs at both edges of both surfaces of a lens, if the lens is thin enough, it is often a good approximation to just assume there is one bend somewhere in the lens.

- There is a convex-lens refraction at the curved surfaces of the lens that sends parallel rays to a focus. If you put a light source at the focus, rays will go out diverging, and the lens will make them parallel—which is another way to make a search light.

- A convex lens does what a concave mirror does: It converges parallel light rays to a focus. A magnifying glass is an example of a convex lens.

- The lens on an old-fashioned film projector is a lens that projects a real image on the screen of the projector. This tells you that the film goes through upside down so that what you see on the screen is right side up.

- By tracing rays with lenses, real images or virtual images are created, depending on where the object is placed relative to the focal point. The real images can be either reduced or enlarged, depending on whether it is outside twice the focal length or inside twice the focal length.

- In Figure 42.1, there is a convex lens that has a focal length f and 2 focal points F, one on either side of the lens. A height h is some distance beyond the focal point, and the distance between the lens and the object (O) is s.

Figure 42.1
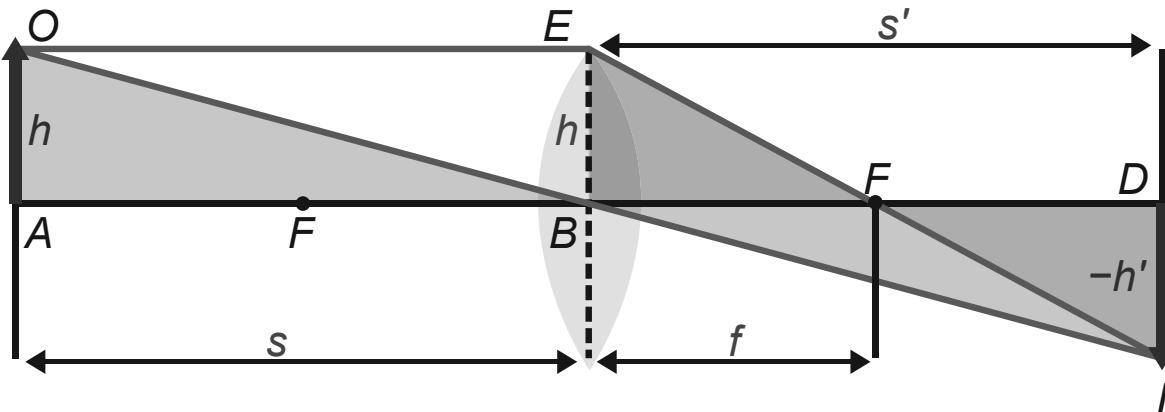
- There are 2 rays entering the lens: one goes parallel to the lens axis and then is bent through the focal point, and the other goes right through the center of the lens and reaches the first ray at some point *I* past the lens. That point defines the image.

- At point *I*, the image height is *h′* (*h* prime). If *h′* is negative, then the image is inverted. If *h′* is positive, the image is not inverted, and the actual length of the arrow on the right-hand side of the figure is −*h′*.

- The distance to the image is *s′*. There are 2 triangles—triangles *OAB* and *IDB*—that are both right triangles and, therefore, are similar.

- Because they are similar triangles, you can write the following ratio: $\dfrac{h'}{h} = -\dfrac{s'}{s}$. This equation defines the magnification *M*.

- The magnification is positive if the image is upright or negative if the image is inverted. In this case, the magnification is negative.

- There are a couple more similar triangles: triangles *EBF* and *IDF*. From these, another ratio can be formed—the ratio of *h*, the height of the object, to *f*, the focal length: $\dfrac{h}{f} = \dfrac{-h'}{(s'-f)}$.

- Algebraically combining the equations for the 2 pairs of similar triangles, you get the lens equation: $\frac{1}{s}+\frac{1}{s'}=\frac{1}{f}$, which does quantitatively with numbers what the geometrical process of ray tracing does with geometrical shapes like lines.

- If s' is greater than zero, the real image is inverted. It is on the opposite side of the lens from the object, which is what the figure shows.

- The lens equation also covers the situation of a virtual image, in which case s' is less than zero, telling us that the image is on the same side of the lens as the object and that it is an upright, virtual image.

- The lens equation is a mathematical way of dealing with how to quantify what goes on with lenses instead of ray tracing.

- Vision correction is a nice application of the lens equation, which is a formula for how images get refracted through a lens.

- In addition to **converging lenses** (convex lenses), which send parallel rays to a focus, **diverging lenses** (concave lenses) send parallel rays away from a focus and can only form virtual images.

- The lens equation still works for diverging lenses, but you have to consider that the focal length of a diverging lens is negative. You have negative image distance s' because the image is on the same side of the lens as the object.

- The problem with a nearsighted eye is that the focus by the natural eye's cornea and lens is not far enough back to be on the retina, and you can correct that by using a diverging lens.

- All that matters for a diverging lens is that it be thinner in the center than at the edges. It could still have a curvature at the front, a positive curvature, if the inner surface is more sharply curved.

Important Terms

concave mirror: This type of curved mirror is a device that reflects light, forming an inverted real image that is in front of the mirror.

converging lens: A type of lens that sends parallel rays to a focus.

convex lens: A lens that bends outward by taking parallel rays and bending them to a focal point.

diverging lens: A type of lens that sends parallel rays away from a focus and can only form virtual images.

lens: A piece of transparent material shaped so that refraction brings light rays to a focus.

real image: An image that is small, inverted, and beyond the focal point (where 2 rays meet) in which the eye is seeing light that is actually coming from the image.

Suggested Reading

Rex and Wolfson, *ECP*, chap 21.4–21.6.

Wolfson, *EUP*, chap 31.

Questions to Consider

1. Can a concave mirror form a real image? A virtual image? Speculate on the answer to this question for a convex mirror.

2. Is the image you see in a magnifying glass real or virtual?

3. In the lens equation, what's the significance of a negative focal length?

Wave Optics

Lecture 43

Light is a wave, and 2 or more light waves undergo interference, enhancing the wave amplitude where wave crests meet and diminishing it where crest and trough meet. Light's wave nature is often demonstrated through the interference of light passing through 2 narrow slits. Extending this to multiple slits gives diffraction gratings that can separate light by wavelength with exquisite precision. Light undergoes diffraction, a bending of its path, when light waves pass by sharp edges or through apertures. Diffraction ultimately limits our ability to image small or distant objects.

- In this lecture, we're going to remember what we intentionally forgot in the previous 2 lectures—the fact that light consists of waves.

- Remember that the approximation of geometrical objects, ray approximation, could be used when the systems we were concerned with, lenses or mirrors, were large compared to the wavelength of light.

- Physical optics—as opposed to geometrical optics or ray optics— becomes important when light interacts with systems that are the size of the wavelength.

- The wave nature of light fundamentally limits our ability to image objects—particularly to image either small objects or distant objects that are very close together.

- The fundamental phenomenon that waves experience but that matter does not experience is the phenomenon of interference, which deals with the wave aspects of light and how they affect optical phenomena.

- Remember that interference can be either constructive, which occurs when wave crests meet each other and wave troughs meet each other, or destructive, which occurs when crests meet troughs.

- Interference in thin films can be used to make very precise optical measurements of very small distances using a method called interferometry. In other words, you can look at interference patterns to infer distances.

- Thomas Young used interference in the early 1800s to prove definitively that light consisted of waves by using 2-source interference, which was reviewed in Lecture 18.

- If you have 2 sources of light and assume they're coherent—that is, they are emitting light waves in exactly the same phase of exactly the same frequency—they will send out circular wave fronts in 3 dimensions that will radiate outward from the source.

- For these 2 sources of light, there are places where wave crests cross wave crests and wave troughs cross wave troughs. These are regions where the waves are interfering constructively.

- In between is a place where a wave crest crosses a trough. As the waves propagate outward, nevertheless we still have these same regions, or bands, in which there is constructive and destructive interference.

- Constructive and destructive interference produce **interference fringes**, which are alternating bright and dark bands.

- In between these regions or bands, there are regions where the waves almost—not quite, but almost—cancel each other out, and that's 2-source interference.

- Instead of trying to make 2 coherent sources of light, you instead put a single source of light through 2 small holes or slits, which results in **2-slit interference** (as opposed to 2-source interference).

- In 2-slit interference, incoming waves go through the 2 slits in some barrier, and each slit acts as a new source of circular waves. As a result, the interference pattern is exactly the same as 2-source interference.

- Figure 43.1 depicts a 2-slit system. The slits are vertical, and the slit spacing is d. There is a screen some distance L away, so L is the distance between the slits and the screen. There are paths from one slit to some arbitrary point P on the screen, which is a distance y away from the center line between the 2 slits.

Figure 43.1
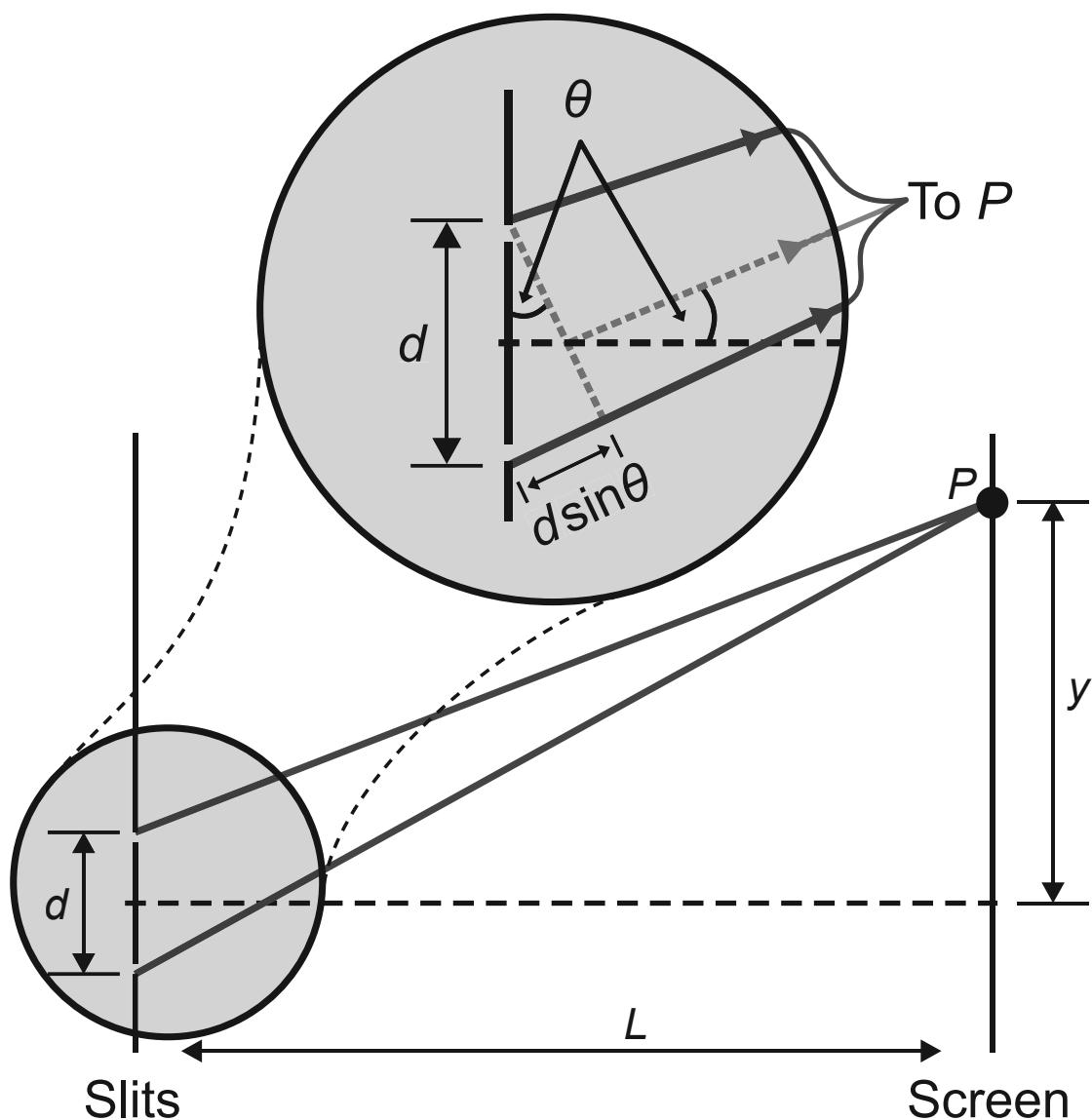
- Where are the bright and dark fringes of this interference pattern? Fringes r are a function of the distance y.

- For constructive interference, the path difference $r_2 - r_1$ is a multiple of the wavelength λ: $d\sin\theta = m\lambda$, where m is the order (number) of the fringe.

- For a small angle θ, the sine of θ is approximately equal to the tangent of θ, which is equivalent to y/L.

- Therefore, bright fringes occur when $d\dfrac{y}{L} = m\lambda$, which is equivalent to $y = m\dfrac{\lambda L}{d}$.

- For constructive interference, the path difference $r_2 - r_1$ is an odd multiple of a half-wavelength. Therefore, dark fringes occur when $y = \left(m + \tfrac{1}{2}\right)\dfrac{\lambda L}{d}$.

- Solving the bright-fringe equation for λ, the equation for the first fringe ($m = 1$) is $\lambda = \dfrac{yd}{L}$.

- From this equation, it turns out that the wavelength of light is 650 nm, which is the color of visible red light. That's a small quantity that is quite accurately measured using 2-slit interference.

- What happens if there are multiple slits? The criterion for a maximum (constructive interference) is still $d\sin(\theta) = m\lambda$.

- The criterion for a minimum (destructive interference) is more complicated: It has to be some integer divided by the total number of slits multiplied by λ: $d\sin\theta = (m/N)\lambda$, where N is the number of slits.

- The integer can't be a multiple of the number of slits, or you'd have constructive interference.

- When the number of slits is relatively small, you get the kind of interference pattern discovered in Figure 43.1.

- As the number of slits increases, the brightness of the fringes increases, and the fringes get increasingly narrow.

- If you have a large number of slits, you have very bright and very narrow fringes in the interference pattern, and in between, you have almost nothing.

- This is ultimately the principle behind a diffraction grating, which is when you have an extremely large number of slits that will reflect light. This can occur either with reflection or slit-like systems.

- If there are a great many of slits, then individual wavelengths will be sent out very precisely at different angles, and you'll see bright regions for different wavelengths of different angles. This can be used to break up light into its colors, like a prism.

NRAO/AUI.

Radio astronomers can obtain extremely precise measurements and imaging with groups of radio telescopes called interferometric telescopes.

- **Diffraction**, the bending of light waves as they pass objects, is best understood by a principle called **Huygens's principle**, which is named for the Dutch physicist Christiaan Huygens.

- Huygens argued that each point on a wave crest can be treated as a source of expanding spherical waves, which then interfere to produce propagating waves.

- Each point on a straight wave makes these circular wave fronts, and they eventually interfere to make the later position of the wave front. If the wave front is circular, then you get an advanced and somewhat larger circular wave front as the wave propagates outward.

- Diffraction occurs as waves go around an edge. Each of the wave fronts produces circular wavelets, called Huygens wavelets, and right near the edge, there's no other waves to interfere, which results in a curvature around the edge.

- Diffraction also takes place at **apertures**, which are any kind of holes that light might pass through—for example, slits and circular holes.

- When light is sent through an aperture, there will be some interference patterns on the other side caused by the interference of light rays coming from different parts of the aperture. This takes into account the finite size of the aperture.

- When the aperture size is comparable to the wavelength, the images become spewed all over the place: Diffraction prevents us from easily forming images.

- To overcome the **diffraction limit**, very large apertures (enormously large telescopes) are needed to image either small or closely spaced, distant objects—or a smaller wavelength needs to be used.

Important Terms

2-slit interference: A process whereby incoming waves of light go through 2 slits in some barrier, and each slit acts as a new source of circular waves. This process results in the same interference pattern as 2-source interference.

aperture: Any kind of hole that light can pass through and at which diffraction can occur.

diffraction: The phenomenon whereby waves change direction as they go around objects.

diffraction limit: A fundamental limitation posed by the wave nature of light, whereby it is impossible to image an object whose size is smaller than the wavelength of the light being used to observe it.

Huygens's principle: Named for the Dutch physicist Christiaan Huygens, this principle explains the process whereby each point on a wave crest can be treated as a source of expanding spherical waves, which then interfere to produce propagating waves.

interference fringe: Alternating bright and dark bands that are produced by constructive and destructive interference.

Suggested Reading

Rex and Wolfson, *ECP*, chap 22.1–22.3.

Wolfson, *EUP*, chap 32.

Questions to Consider

1. Under what conditions does the wave nature of light become important in describing optical phenomena?

2. Why do you see colors in a soap film or an oil slick?

3. How do multiple-slit systems enable the precise separation of light by wavelength?

Cracks in the Classical Picture
Lecture 44

B y the late 19th century, physicists thought they had the physical world thoroughly explained, but a few unexplained cracks in the classical picture led to radically new physics in the 20th century. Ether, which presumably permeated the entire universe, proved difficult to detect. In the solar system, the planet Mercury presented a small but unexplained deviation from Newton's gravitational theory. On another front, the classical description of radiation from hot, glowing objects was in dramatic contradiction with observations. So was the interaction of newly discovered electrons with light. Furthermore, Maxwell's electromagnetism suggested that atoms shouldn't exist.

- Classical physics, by the late 19th century, was a physics in which mechanics is well understood. Electromagnetism, which was summarized by Maxwell's equations, now attempted to be understood as a mechanical phenomenon.

- This mechanical view is the view that was swept away in several big revolutions, including the relativity revolution and the quantum revolution of the early 20th century.

- There are several cracks in classical physics. Some of these cracks are extreme disconnects between what theoretical physics predicts and what observation or experimentation shows, but some of these cracks are more subtle.

  - The first crack focuses on the quandaries that surround the substance ether.

  - In addition, there are quandaries involving Mercury's orbit.

  - There are problems involving glowing objects—hot objects that glow hot enough to give off light.

○ There are also cracks involving the interaction between electrons and light.

○ Finally, there are cracks in classical physics that involve the very nature and, in fact, the very existence of atoms.

- Of these 5 quandaries, all but the one involving Mercury's orbit, which is a gravitational quandary, have something to do with electromagnetism.

- First of all, ether, which Maxwell and others came up with, was supposed to be a medium in which electric and magnetic fields were stresses and strains and, therefore, in which electromagnetic waves were propagating disturbances.

- Ether is an improbable substance. It would have to permeate the entire universe because electromagnetic waves permeate the entire universe.

- It would also have to be a very stiff substance because the waves propagate at the very high speed of light. Ether has to be stiff, and yet the planets and stars have to be able to move through it essentially without resistance.

- Lecture 18 discussed the fact that electromagnetic waves are transverse waves. The electric and magnetic fields are at right angles to the direction that wave is moving, and that means the ether couldn't be a gas like air.

- It also couldn't be a liquid like water because those kind of materials don't support transverse waves. Therefore, it had to be more like Jell-O, which made it even more improbable.

- Maxwell's equations talk about electric and magnetic fields as disturbances of the ether, so his equations are valid in a reference frame that's fixed with respect to the ether.

- Therefore, deviations can be detected from Maxwell's equations if using reference frames that are moving with respect to this mysterious ether substance.

- Earth's motion relative to the ether has to be changing, so differences in the speed of light in different directions should be detectable, but the quandary is that they aren't.

- The second of the cracks deals with Mercury's orbit. In classical Newtonian gravitation, the prediction is that a planet ought to undergo an elliptical orbit.

- In fact, Mercury's orbit does not close on itself, but the whole plane, the whole axis of the orbit, moves slowly with time—a phenomenon called **precession**.

- Some of that we understand is due to the presence of other planets and other effects that are explained by Newtonian physics. When you remove those explanations, however, there's still an unexplained precession of 43 seconds of angle.

- Moving on to the third crack, when you apply a variable electric current to an ordinary light bulb, it begins to glow (not very hot). The atoms in the filament of that light bulb are being jostled into more rapid thermal energy.

- As you turn the current up, the light bulb becomes brighter because its temperature is higher; the atoms are jostling around more vigorously because temperature is a measure of the average kinetic energy.

- As you increase the temperature of the light bulb, the energy shifts toward shorter wavelengths—from dull red to orange to a yellow color.

- Classical physics predicts that the energy from a glowing object should rise continuously (producing infinite energy) with decreasing wavelength. This notion is in complete contrast to the observation, which is if the energy peaks at a wavelength, it depends on temperature.

- Because the shortest wavelengths known at the time of classical physics were the wavelengths associated with ultraviolet radiation, this was called the **ultraviolet catastrophe**.

- The fourth crack involves the behavior of electrons in light through a phenomenon called the **photoelectric effect**, which had been studied in the 1880s when people were beginning to experiment with evacuated tubes.

- Evacuated tubes were the precursors of the vacuum tubes, which powered electronics through the first half of the 20th century and then were replaced by transistors in solid-state electronics.

- Classical physics predicted that in an evacuated tube, it would take a significant amount of energy for an electron to be ejected from a metal and that increasing the intensity of the light that is shining on the metal should shorten that time because the wave gets bigger and so do the electric fields.

- Classical physics also predicted that the electron energy shouldn't depend on the light's wavelength. However, when you do the experiment, you find that increasing light intensity doesn't shorten the time and that the electron energy actually does depend on wavelength.

- Electron energy increases with decreasing wavelength, so a shorter wavelength has a higher frequency, and there's actually a cutoff wavelength at which point no electrons are ejected.

- The last of the cracks in classical physics involve atoms. In around 1911, after a series of experiments, Ernest Rutherford and several colleagues developed the idea of the nuclear atom—with electrons orbiting around the nucleus.

- The orbiting electrons are accelerated charges because they're going in circular motion, in which change in direction matters as much as change in speed.

- According the classical physics, these accelerated charges ought to radiate electromagnetic waves. The electrons should lose their kinetic energy of their motion and spiral into the nucleus quickly—and atoms shouldn't exist. That's a stark contrast to what happens because atoms do exist.

- Finally, from this idea of classical physics, atoms should emit a continuous range of wavelengths or colors—visible light—as the electrons spiral into the nucleus, but you don't see a whole rainbow of colors coming from an atom.

- Instead, atoms radiate discrete wavelengths, called spectral lines, and the different patterns of these spectral lines are characteristic of each element.

Important Terms

photoelectric effect: A phenomenon in which light incident on a metal surface causes electrons to be ejected from the surface. The analysis of the photoelectric effect by Albert Einstein was an early success of quantum theory.

precession: The gradual change in direction of a rotating object's rotation axis as a result of an applied torque.

ultraviolet catastrophe: The absurd prediction of classical physics that a hot, glowing object should emit an infinite amount of energy in the short-wavelength region of the electromagnetic spectrum.

Questions to Consider

1. How did the success of Newtonian physics incline 19th-century physicists toward a mechanical view of light waves?

2. Describe 3 of the cracks in classical physics that opened the way for the new physics of the 20th century.

Earth, Ether, Light
Lecture 45

B ased on their view that light should be a mechanical wave, 19th-century physicists sought to find the medium—the ether—of which light was a disturbance. These physicists, basing their assumptions on Maxwell's equations and the speed of light, asked a logical question: What is Earth's motion through the ether? Using an exquisitely sensitive apparatus based on the interference of light, the Michelson-Morley experiment nevertheless failed to detect any motion of Earth relative to the ether. This result led to a deep-seated contradiction at the heart of physics.

- Ether is a very improbable substance. Because Earth's motion can't be detected relative to it, maybe we should just dispense with it. If we do that, however, we have to ask what the medium for light is—in what medium light is a disturbance.

- Ether allowed us to understand electromagnetism and electromagnetic waves like we would any kind of mechanical waves absent the ether.

- With respect to what does light travel at speed c? The answer is obvious if we have ether: It travels at speed c with respect to the ether.

- Without the ether, maybe light travels at speed c with respect to its source, but it turns out that it doesn't.

- Observations of binary stars, a system in which 2 stars are in gravitational orbits around each other, are the perfect example that light is not traveling at speed c relative to its source.

- If light moved at speed c relative to its source, then when the star was moving toward you, the light would be coming to you, relative to you, faster than it would be when the star was moving away from you half an orbit later.

- This notion disturbs the picture of the star having a simple circular orbit, and the fact that we spectroscopically see a simple orbit tells us that the speed of light is not determined by the speed of the source.

- Another explanation of the fact that we can't seem to detect Earth's motion through the ether is that maybe Earth is not moving through the ether. Maybe Earth is at rest relative to the ether.

- Other planets are moving relative to Earth, and other stars are moving relative to Earth and to the Sun, so Earth is not at the center of things. Therefore, Earth couldn't be at rest with respect to the ether and have nothing else at rest with respect to the ether.

- Maybe, on the other hand, Earth drags the ether in its vicinity with it. This could work, and each planet would measure a speed of light that was the same and didn't vary with position. However, this is ruled out by a phenomenon called the **aberration of starlight**.

- The aberration of starlight was first measured in 1729 by the English astronomer James Bradley. This phenomenon is similar to trying to walk through rain with an umbrella; even if the rain is falling vertically, you have to tilt the umbrella to stay driest.

- If Earth is moving relative to the ether, it's not dragging the ether with it. The apparent position of a star seems to wobble over time— there is a kind of sinusoidal variation in its position.

- Consequently, we rule out the possibility that Earth is at rest with respect to the ether; therefore, Earth must be moving relative to the ether.

- If Earth is moving through the ether, then there must be a wind of ether blowing past Earth that affects the speed of light in different directions as Earth moves around in its orbit. Therefore, different speeds of light can be measured in different directions.

- Conducted in 1887, the **Michelson-Morley experiment** consisted of sending light on 2 round-trip paths, one of them parallel to the ether wind and one of them perpendicular to the ether wind; they detected subtle differences in travel time.

- The ether wind is blowing with some speed v, and that's the speed of Earth relative to the ether at a given time.

- First, measure a light path that is moving upstream against the ether wind. This light is struggling against that ether wind, and the light speed relative to the ether is c, the speed of light, because the ether is the medium in which light is a disturbance.

- The ether is slowing the light down because the ether wind is blowing against light's motion, so the light speed relative to the ether is $c - v$.

- Therefore, when moving upstream, the light takes time $t_{\text{upstream}} = \dfrac{L}{c-v}$, where L is the length of the path.

- Then, when moving downstream, the ether wind is traveling with the light, so the light is sped up. The speed of light relative to Earth becomes $c + v$, and the light takes time $t_{\text{downstream}} = \dfrac{L}{c+v}$.

- The total round-trip travel time perpendicular to the ether wind is $t_{\text{perpendicular}} = \dfrac{2L}{u} = \dfrac{2L}{\sqrt{c^2 - v^2}}$.

- The total round-trip travel time parallel to the ether wind is $t_{\text{parallel}} = \dfrac{L}{c-v} + \dfrac{L}{c+v} = \dfrac{2cL}{\sqrt{c^2 - v^2}}$.

- The Michelson-Morley experiment shows that Earth's motion relative to the ether is not detectable. In other words, the results fail to detect Earth's motion through the ether.

- Ether is needed to give a mechanical interpretation of electromagnetism. More significantly, ether is needed because the aberration of starlight and the binary star systems tell us that light does not move at speed *c* relative to its source and that Earth is not at rest with respect to the ether.

- Where do things stand now in physics? In 1887, there was a seeming contradiction: Earth is not at rest with respect to the ether, but we can't detect Earth's motion with respect to the ether.

- There was also a philosophical dichotomy: Newtonian mechanics works perfectly in any moving frame of reference, but in the mechanical view of 19th-century physicists, electromagnetism has a preferred reference frame, which is at rest relative to the ether, that we can't detect.

- The **Lorentz-Fitzgerald contraction**, proposed independently by the Irish physicist Fitzgerald and the Dutch physicist Lorentz, states that when objects move through ether, they are compressed in the direction of motion.

- The contraction effect described in this hypothesis explains why you would not be able to detect motion through the ether with an experiment like Michelson-Morley's.

- There turns out to be a grain of truth in the Lorentz-Fitzgerald contraction, but it is based on an ad hoc assumption and clings to the idea that there is this mysterious and problematic substance—ether—that travels through the entire universe.

Important Terms

aberration of starlight: A phenomenon whereby a telescope must be pointed in slightly different directions at different times of year because of Earth's orbital motion. The fact of aberration shows that Earth cannot drag with it the ether in its immediate vicinity and, thus, helps dispel the notion that ether exists.

Lorentz-Fitzgerald contraction: Proposed independently by the Irish physicist Fitzgerald and the Dutch physicist Lorentz, this hypothesis states that when objects move through ether, they are compressed in the direction of motion.

Michelson-Morley experiment: An experiment performed in the late 19th century by the 2 American physicists from which it takes its name, with the goal of detecting the presence of the ether through which electromagnetic waves moved. Its failure to detect any evidence for the ether led to the development of Einstein's theory of relativity.

Suggested Reading

Rex and Wolfson, *ECP*, chap 20.3.

Wolfson, *EUP*, chap 33.1–33.2.

Questions to Consider

1. How does the idea of Earth at rest with respect to the ether violate the Copernican paradigm?

2. Did the Michelson-Morley experiment actually measure the speed of light in 2 different directions? Explain.

3. How did the Michelson-Morley result challenge prevailing assumptions about the propagation of light?

Special Relativity
Lecture 46

Albert Einstein's special theory of relativity states that the laws of physics are the same for any observer in uniform motion. This theory necessitates a radical revision of our notions of space and time because relativity requires that Maxwell's description of electromagnetic waves as traveling at the speed of light in vacuum be valid for all observers—including those in relative motion. That can only be true if different observers have different measures of space and time. Therefore, relativity requires the modification of many Newtonian concepts—but not of electromagnetic concepts.

- The purpose of this lecture is to introduce one simple idea that is going to overturn our ideas of space and time and resolve the ether quandaries previously studied.

- At the end of the 19th century, all attempts to detect Earth's motion through ether, the mysterious, problematic substance that pervaded the whole universe, failed. Instead, the following question was left: With respect to what does light move at speed *c*?

- Centuries of commonsense ideas about space and time hinge on that question, which was resolved by a single individual, who was probably the most brilliant physicist of all time: Albert Einstein.

- Albert Einstein was born in 1879 in Ulm, Germany. The year 1905, during which he was 26 years old, was a miracle year for him. He published seminal papers in 3 very distinct areas of physics, one of which could've won him the Nobel Prize.

- On April 30, he published his doctoral dissertation, and on June 9, he published a paper that is the predominant one that won him the Nobel Prize. This paper introduced one of the key concepts in quantum physics and was entitled "On a Heuristic Viewpoint Concerning the Production and Transformation of Light."

- On July 18, he published convincing proof that atoms exist in his paper entitled "On the Movement of Small Particles Suspended in Stationary Liquids Required by the Molecular-Kinetic Theory of Heat."

- On September 26, Einstein published the paper that describes to the world, for the first time, the special theory of relativity called "On the Electrodynamics of Moving Bodies."

- On November 21, he published what's basically a footnote to special relativity; it's the paper in which the idea $E = mc^2$, probably the most famous equation in the world, was first introduced. The title of the paper is "Does the Inertia of a Body Depend Upon Its Energy Content?"

- At the heart of the paper entitled "On the Electrodynamics of Moving Bodies" is that one simple idea that changes everything. It's called the **principle of relativity**.

- Galileo and Newton knew of the principle of relativity, but Einstein generalized that principle to become what's called the special theory of relativity.

- The **special theory of relativity** states that the laws of physics are the same for all observers in uniform motion. If you remove the phrase "in uniform motion," this statement basically becomes the **general theory of relativity**.

- The special theory of relativity means that no matter who you are, where you are in the universe, or how you're moving, as long as your motion is uniform—unaccelerated, not changing in speed or direction—then you could perform physics experiments and would get the same results for the laws of physics as anyone else who is in uniform motion.

- Earth is close enough to uniform motion that we can talk about Earth as if it were a reference frame in uniform motion, a so-called inertial reference frame.

- In his paper entitled "On the Electrodynamics of Moving Bodies," Einstein referenced the failure to detect Earth's motion through the ether. Historians of science still debate whether Einstein knew of the Michelson-Morley result.

- Einstein also states, for the first time, that it's not just Newtonian mechanics that doesn't depend on your reference frame—it's also electromagnetism. The implications for electromagnetism are going to lead us deeply into what relativity really tells us about the world.

- An alternate way to describe the special theory of relativity is that the phenomena of electrodynamics as well as mechanics possess no properties corresponding to the idea of absolute rest; these phenomena have no preferred frame of reference.

- That ether doesn't exist is part of the implications of relativity. Concepts of absolute rest and absolute motion are meaningless; in other words, only relative motion matters.

- One implication of the special theory of relativity is that this statement—the laws of physics are the same for all observers in uniform motion—applies to all of physics, not just mechanics or electromagnetism.

- Before Einstein's special theory of relativity, everyone thought that there was only one frame of reference—the ether frame—for which electromagnetism was valid and for which Maxwell's equations would give you the correct results where you would measure c for the speed of light.

- The principle of relativity, then, tells us that Maxwell's equations are valid in any frame of reference, and that means that the speed of light is the same for all observers in uniform motion, even if they're moving relative to each other.

- Einstein actually presented 2 principles, which he called postulates, in his paper: One was the principle of relativity, and the other was what he said was a seemingly contradictory statement—that the speed of light was independent of its source.

- Philosophically, however, you do not need both of those postulates. The principle of relativity covers it all because it says that Maxwell's equations have no preferred frame—that they're equally valid in any uniformly moving reference frame.

- Nobody has the right frame of reference, and yet they all measure the speed of light as the same value, and the only way that can happen is if their measures of space and time are different.

- Therefore, space and time cannot be absolutes; they are not what we thought they were with common sense. There is no immutable time flowing that is unaffected by anything throughout the whole universe.

- Newton actually made a statement to the effect that time was an absolute, but he was wrong. The concepts of absolute space and time simply cannot be consistent with the statement that the laws of physics must be the same for all observers in uniform motion.

- Maxwell's electromagnetism, it turns out, is relativistically correct. However, some small corrections are needed in the realm of Newtonian mechanics. Although relativity already worked in mechanics, it didn't use Einstein's special theory of relativity.

- Galilean velocity addition needs to be replaced with relativistic velocity addition, which includes a correction term. Relativistic velocity addition, $u = \dfrac{u' + v}{1 + u'v/c^2}$, precludes traveling faster than c.

Important Terms

principle of relativity: A statement that only relative motion is significant. The principle of Galilean relativity is a special case, applicable only to the laws of motion. Einstein's principle of special relativity covers all of physics but is limited to the case of uniform motion.

special theory of relativity: Einstein's statement that the laws of physics are the same for all observers in uniform motion.

theory of general relativity: Einstein's generalization of special relativity that makes all observers, whatever their states of motion, essentially equivalent. Because of the equivalence principle, general relativity is necessarily a theory about gravity.

Suggested Reading

Rex and Wolfson, *ECP*, chap 20.4–20.5.

Wolfson, *EUP*, chap 33.3–33.4.

Lecture 46: Special Relativity

Questions to Consider

1. Name 2 of the seminal contributions Einstein made to physics in 1905.

2. Articulate the principle of relativity in a single sentence.

3. Did special relativity require modification of Maxwell's theory of electromagnetism? Explain.

Time and Space
Lecture 47

I n relativity, measures of time and space are not absolute; instead, they depend on the observer. Furthermore, in some cases, different observers may disagree even about the time order of 2 events. But the apparent contradiction here is resolved by the fact that, in relativity, events that are simultaneous in one reference frame aren't simultaneous in another reference frame moving relative to the first. The different measures of time and space in different reference frames are evident in the famous twin paradox.

- This lecture is an examination of the implications for the nature of space and time of the essential principle of relativity. We want to understand quantitatively how it is that our commonsense notions of space and time simply aren't right—we have to modify them.

- One of the key concepts in relativity is the concept of an **event**, which is an occurrence that you specify by giving both its place and its time.

- Mathematically, an event is specified by the existence of 4 numbers—3 spatial coordinates on an arbitrarily chosen *x*, *y*, and *z* coordinate system and a time coordinate, which tells us when the event occurs relative to some arbitrary zero of time.

- All observers agree about the reality of events, but what they may disagree about are the space and time coordinates of an event and about the time interval and spatial distance between 2 events.

- This lecture focuses on the time interval and the spatial distance between 2 events, which cannot be the same for all observers—a statement that is not in contradiction with the principle of relativity.

- The principle of relativity does not say that all observers will measure the same values for all things. What it does say is that all observers will come up with the same general laws of physics, including the modification of Newtonian mechanics and Maxwell's equations of electromagnetism (which are already relativistically correct).

- Notions of space and time—measures of spatial distances and temporal intervals—cannot be the same for all observers.

- Imagine a device called a light clock, which is just a timekeeping mechanism. Like any other timekeeping mechanism, it has some periodic occurrence that sets the timekeeping of the clock.

- In this clock, the timekeeping is set by a light source that flashes, sends a light beam up to a mirror, and then heads back down to the light source.

- The ticking of the clock consists of the interval between the light leaving the source and the light returning to the source. As shown in Figure 47.1, the light clock has a length L in its long dimension.

Figure 47.1
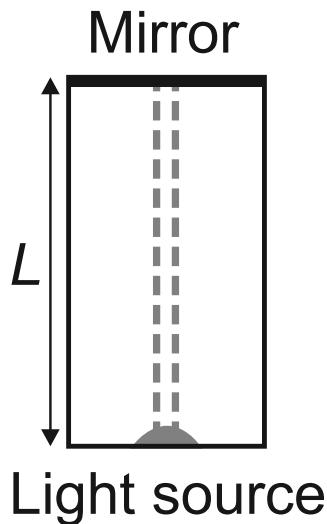
- Even though this device is a rather specialized construct that involves the properties of light, the conclusions drawn about this specific device are going to be conclusions about any clock because they're ultimately conclusions about time itself.

- Let's define 2 events: Event A is the light flash leaving the source and heading up toward the mirror, and event B occurs when the light returns to the source.

- What's the time interval between event A and event B, given that the length of the light clock is L? Assuming these events are occurring in a vacuum, the speed of light is c.

- In general, we actually have to begin by asking what the reference frame is because many quantities in physics that we think of as being absolutes depend, in fact, on your point of view—or frame of reference.

- The reference frame in which the light clock is at rest is S'. In that frame of reference, the light travels a round-trip distance $2L$ up and back. It travels at the speed of light, which is c. Distance = speed × time, so $\Delta t' = \dfrac{2L}{c}$.

- In a reference frame S, as depicted in Figure 47.2, the light clock moves to the right at speed v. Now you can consider yourself to be in this frame, in which we ask the same question regarding the time between the same 2 events.

- If we didn't believe relativity, we would end up with the same answer for reference frame S as we did for reference frame S', but we're not, and that's going to be our first insight into the difference between time in classical physics and time in relativity.

Figure 47.2
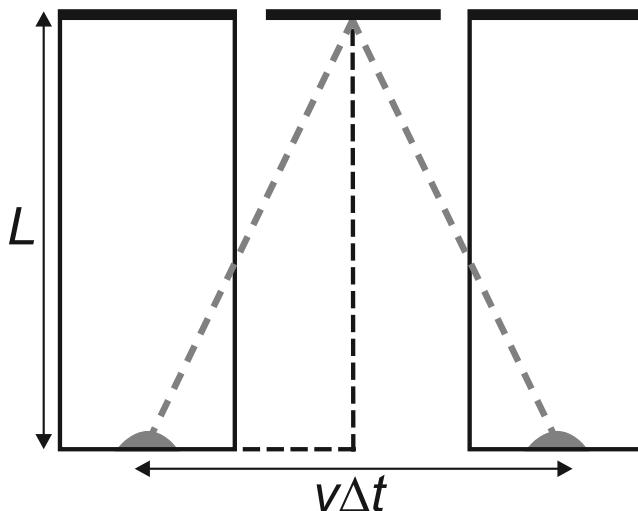
- In this reference frame, the total distance is twice the diagonal path, and the light clock is moving at speed v. Using the Pythagorean theorem to find the hypotenuse of the triangle that is formed by the light's path, the total distance becomes $2\sqrt{L^2 + (v\Delta t/2)^2}$.

- According to the principle of relativity, you can't add to the speed of light; it is c in all reference frames, so we can assume that light speed is c.

- Distance is speed times time, and consequently, time is distance over speed, so $\Delta t = \dfrac{2\sqrt{L^2 + (v\Delta t/2)^2}}{c}$ is the distance traveled in this reference frame.

- This is not our answer, though, because there are 2 Δt's in the equation, and we want to solve for Δt so that it only shows up on the left-hand side of the equation.

- Algebraically, $\Delta t = \dfrac{2\sqrt{L^2 + (v\Delta t/2)^2}}{c}$

$$\Rightarrow \quad c\Delta t = 2\sqrt{L^2 + (v\Delta t/2)^2}$$

$$\Rightarrow \quad c^2\Delta t^2 = 4\left(L^2 + (v\Delta t/2)^2\right)$$

$$\Rightarrow \quad c^2\Delta t^2 = 4\left(L^2 + v^2\Delta t^2/4\right)$$

$$\Rightarrow \quad c^2\Delta t^2 = 4L^2 + v^2\Delta t^2$$

$$\Rightarrow \quad c^2\Delta t^2 - v^2\Delta t^2 = 4L^2$$

$$\Rightarrow \quad \left(c^2 - v^2\right)\Delta t^2 = 4L^2$$

$$\Rightarrow \quad \left(1 - v^2/c^2\right)\Delta t^2 = \dfrac{4L^2}{c^2}$$

$$\Rightarrow \quad \dfrac{2L}{c} = \sqrt{1 - v^2/c^2}\;\Delta t \,.$$

- For the frame of reference S', substitute Δt for $\dfrac{2L}{c}$ to get $\Delta t' = \sqrt{1 - v^2/c^2}\;\Delta t$ or $\Delta t' = \Delta t\sqrt{1 - v^2/c^2}$.

- If the velocity were zero, you would be left with the square root of 1, which is 1, and the time interval between the 2 events would be the same for both frames of reference. However, if the v is not zero—if there is relative motion between the 2 frames—then they will measure different values for the time interval between those 2 events. This phenomenon is called **time dilation**.

- In special relativity, time dilation is the phenomenon whereby the time measured by a uniformly moving clock present at 2 events is shorter than that measured by separate clocks located at the 2 events.

- In general relativity, time dilation is the phenomenon of time running slower in a region of stronger gravity (greater space-time curvature).

- The time interval between 2 events is shortest in a reference frame where they occur at the same place. This idea also leads to length contraction, which says that an object is longest in its own rest frame.

- One of the biggest ideas about time and relativity is that simultaneity, the idea that 2 events occur at the same time, is relative. If measures of time are different in different reference frames, so is the measure of simultaneity.

- The different measures of time and space in different reference frames are evident in the famous twin paradox, in which one twin ventures out on a high-speed roundtrip and returns to find herself younger than her Earthbound twin.

- Thus, special relativity allows time travel into the future—although not into the past. In special relativity, motion itself becomes meaningless in an absolute sense—only relative motion matters.

- However, change in motion is still meaningful, and the reason that the 2 twins can experience different aging is because the traveling twin turns around—undergoes a change in motion—while the stay-at-home twin doesn't.

Important Terms

event: A point in space-time, designated by its location in both space and time.

time dilation: In special relativity, the phenomenon whereby the time measured by a uniformly moving clock present at 2 events is shorter than that measured by separate clocks located at the 2 events. In general relativity, the phenomenon of time running slower in a region of stronger gravity (greater space-time curvature).

Suggested Reading

Wolfson, *EUP*, chap 33.5–33.6.

Questions to Consider

1. Do moving clocks run slow? Discuss.

2. A spacecraft goes whizzing by Earth at close to the speed of light. You claim it's 10 meters long; the pilot claims it's 50 meters long. Who's right?

Space-Time and Mass-Energy
Lecture 48

R elativity certainly alters our commonsense notions of space and time, but contrary to popular belief, the theory of relativity doesn't say that everything is relative. What's important is what's not relative—that is, what doesn't depend on one's frame of reference. In relativity, space and time merge into a 4-dimensional space-time in which events are points specified by their space and time coordinates, and intervals between events are 4-dimensional vectors. Mass and energy similarly join, related by Einstein's famous $E = mc^2$ and other equations.

- A very common philosophical view, and misunderstanding of the theory of relativity, is that relativity plays around with commonsense ideas of space and time and makes everything relative.

- In the early 1900s, people were using Einstein's theory of relativity to justify relativistic morals and relativistic aesthetics—all kinds of fields far from science.

- Physics is about trying to understand an underlying objective reality, and that reality should not depend on one's point of view or frame of reference.

- This lecture is about identifying such quantities and developing from them the important ideas of 4 dimensions and 4-dimensional space-time.

- The speed of light is one quantity that doesn't depend on your point of view, but more fundamentally, the laws of physics do not depend on your point of view.

- The numbers obtained when using the laws of physics, at least for some quantities—such as spatial differences or temporal intervals—do depend on your point of view. Other numbers don't—such as, for example, the speed of light.

- The difference between the square of the time interval between 2 events and the square of the spatial interval, or the distance, between the 2 events is $\Delta t^2 - \Delta x^2$.

- The quantity $\Delta t^2 - \Delta x^2$ is the same for different reference frames. It is a quantity that is objectively real and does not depend on your frame of reference. It is a quantity that is related to both space and time, which is why it is called **space-time**.

- Space-time is **invariant**, which means that it doesn't depend on your point of view; it doesn't change with your frame of reference.

- The individual measures of space and time are different in the 2 different frames of reference, but you can combine them to get something that is objectively real: the invariant space-time interval.

- The invariant space-time interval is equal to the difference between the temporal and spatial intervals: $\Delta s^2 = \Delta t^2 - \Delta x^2$. These quantities are equal in any frame of reference in uniform motion.

- It might bother you that events simultaneously in one frame aren't simultaneous in another and that there are different time intervals between different events. However, neither of these concerns turns out to be a problem for causality.

- The space-time interval is analogous to a vector in 2 dimensions but is called a **4-vector** because it is a vector quantity in 4-dimensional space-time that has 4 components: 1 time component, as does the space-time interval, and 3 space components.

Lecture 48: Space-Time and Mass-Energy

- The magnitude of a 4-vector is invariant, and the square of the magnitude is the square of the time component minus the square of the space component.

- The paradigm 4-vector has a time component $c\Delta t$ and 3 space components Δx, Δy, and Δz, but there are many other 4-vectors.

- The **momenergy 4-vector** is a vector that combines energy and momentum into one 4-dimensional mathematical vector. Its time component is an object's total energy, heat; its 3 space components are 3 components of momentum.

- Remember that momentum, from Newtonian physics, is $p = mv$. This equation is slightly modified in relativity to yield the momenergy 4-vector: $\mathbf{p} = \dfrac{m\mathbf{v}}{\sqrt{1 - v^2/c^2}} = \gamma m\mathbf{v}$.

- The total energy of an object is the sum of the kinetic energy and rest energy: $E = \gamma mc^2 = \dfrac{mc^2}{\sqrt{1 - v^2/c^2}}$.

- The invariant magnitude of the momenergy 4-vector is proportional to the object's mass: $E^2 = p^2c^2 + (mc^2)^2$, in which E is the total relativistic energy.

- This statement is called the relativistic energy-momentum relationship, but it is really a statement about the magnitude of the mass of the object.

- This statement shows that, in relativity, the mass of an object is invariant; in other words, it is a property of an object and does not depend on frame of reference.

- An object's energy and momentum do depend on the frame of reference, but they depend in such a way that they conspire to give the same invariant magnitude to the momenergy 4-vector, and that invariant magnitude is essentially within a constant—the object's mass.

- If an object is at rest, p is zero; it has no momentum in its rest frame. In an object's rest frame, the equation $E^2 = p^2c^2 + (mc^2)^2$ becomes $E = mc^2$, Einstein's famous equation, after taking the square root of both sides.

- Therefore, when you're in the rest frame of an object, the relationship between energy, momentum, and mass reduces to $E = mc^2$. This tells you that an object has an energy equal to mass times the square of the speed of light even when the object is at rest.

- $E = mc^2$ is one of the most famous equations in all of physics, and it states that there is energy contained in an object even when it's at rest. This equation expresses the interchangeability of matter and energy, and most importantly, it applies to all energy in matter.

© Comstock/Thinkstock.

The Sun converts about 4 million tons of its mass to energy each second, spewing out energy at the rate of about 4 × 10^{26} watts.

- This equation is commonly, but wrongly, associated only with nuclear energy, but there is nothing nuclear about $E = mc^2$. The discrepancy in mass that is created by releasing energy is measureable in the nuclear case, but it's immeasurably small in the chemical case.

- If light consisted of massless particles, then *m* would be zero and $E^2 - p^2c^2 = 0$. By taking the square root of both sides of that equation, you are left with $E = pc$ or $p = E/c$ for the relationship between energy and momentum of a light particle.

- The momenergy 4-vector is not the only kind of 4-vector—there are others. Electric charge and current, wavelength and frequency, and a host of other quantities assume a new reality in 4-dimensional space-time.

Important Terms

4-vector: A vector quantity in 4-dimensional space-time that has 4 components: 1 time component and 3 space components.

invariant: A quantity that has a value that is the same in all frames of reference. The space-time interval is one example of a relativistic invariant.

momenergy 4-vector: A vector that combines energy and momentum into one 4-dimensional mathematical vector. Its time component is an object's total energy, heat; its 3 space components are 3 components of momentum.

space-time: The unification of space and time required by Einstein's theory of relativity as a consequence of the finite and immutable speed of light.

Suggested Reading

Rex and Wolfson, *ECP*, chap 20.6.

Wolfson, *EUP*, chap 33.7.

Questions to Consider

1. Name 2 quantities whose values don't depend on one's frame of reference. Name another 2 quantities that do depend on reference frame.

2. $E = mc^2$ is often invoked to explain nuclear energy. Is this correct? Does it capture the full meaning of the equation?

3. What is momenergy?

General Relativity
Lecture 49

E instein's special relativity prohibits information from traveling faster than light. Newtonian gravitation is, therefore, incompatible with special relativity. Furthermore, special relativity is limited to reference frames in uniform motion. After special relativity, Einstein turned his attention to gravity. His happiest thought, in 1907, was the realization that gravity and acceleration are locally indistinguishable. This allowed him to generalize relativity to include accelerated frames of reference and incorporate gravity. The result was Einstein's theory of general relativity, in which gravity is not a force but the geometrical structure of space-time.

- Because Einstein's special theory of relativity is limited to the case of observers in uniform motion, a theory of general relativity is needed. We would like to generalize special relativity to remove that limitation, but there are more precise motivations for general relativity that drove Einstein.

- Special relativity talks about uniformly moving—or inertial reference frames, where the law of inertia is obeyed—in which if an object is subject to no forces, it stays in uniform motion.

- It becomes difficult to identify uniformly moving frames of reference; you can identify nongravitational forces, but identifying gravitational forces is very difficult.

- Special relativity is a theory of instantaneous action and distance, and instantaneous transmission of any kind of influence would violate causality under special relativity because it would require an influence to travel faster than the speed of light—which it can't.

- Special relativity is not compatible with Newtonian gravitation, so gravitation has to be modified to make it compatible with general relativity. For that reason, general relativity becomes primarily a theory about gravity.

- Einstein's so-called happiest thought occurred in 1907, and it was the idea of the principle of equivalence, which says that the effects of gravity and acceleration cannot be distinguished.

- This notion states that we can't tell for sure whether we're in a uniformly moving frame of reference with no forces acting or whether there are forces acting and we're not in a uniformly moving frame of reference.

NASA, ESA, K. Sharon (Tel Aviv University) and E. Ofek (Caltech).

NASA's Hubble Space Telescope captured the first picture of a group of 5 starlike images of a single distant quasar. The multiple-image effect is produced by gravitational lensing.

- Over vast expanses of space-time or in regions where gravity is extremely strong, the principle of equivalence is not true, and that's essentially the global reach of general relativity.

- In small enough regions of space-time—limited in space and time—then you can't distinguish the effects of gravity from acceleration. That's why it's hard to identify inertial reference frames.

- In an object that is in freefall, all other objects around it are also in freefall—experiencing the same acceleration—so it appears that gravity is not there. It appears that gravity is transformed away by entering freefall.

- The only things that are real are the things that don't depend on point of view. That means that if there's a frame of reference in which gravity doesn't seem to be, then what is thought of as gravity—that massive force that pulls objects to Earth—is, in some sense, not real.

- The next philosophical leap is to determine that all reference frames are equivalent—not just uniformly moving ones. If we make that leap, then when we're in freefall, we're in a frame of reference where gravity has been transformed away. What is gravity, then, if it isn't a large, remote force of the mass of Earth that pulls objects downward?

- The notion of gravity can be understood by looking at a larger region of space-time. In doing so, you'll notice differences in the direction of gravity.

- The forces that are caused by differences in gravity from place to place are called **tidal forces** because they're what also cause the tides. Contrary to what many believe, the tides are not caused by gravity.

- In Einstein's view, tidal forces are the essence of gravity because, unlike gravity, tidal forces cannot be transformed away by jumping into another reference frame. Furthermore, things that are independent of reference frames are objectively real.

- Einstein says that these tidal forces are the geometry of space and time—the geometry of 4-dimensional space-time. This geometry can be curved, and objects can move in strange ways within this curved geometry, and that's how Einstein explains gravity.

- The essence of general relativity is summarized in 2 statements: The presence of matter and energy changes the geometry of space-time to be a curved space-time, and objects not subjected to forces move in the straightest possible paths in curved space-time.

- The first statement defines the origin of gravity, which is not a force; instead, gravity is the curvature of space-time.

- The equation that explains how the geometry of space-time arises from the presence of mass and energy is

$$R_{\mu\nu} - (1/2)g_{\mu\nu}R + g_{\mu\nu}\Lambda = (8\pi G/c^4)T_{\mu\nu}.$$

- This equation states that at a point in the universe where the mass-energy is described by $T_{\mu\nu}$, the geometry of space-time is described by G. It is the description of gravity in general relativity.

- One of the many applications of the notion that matter moves on the straightest possible paths in curved space-time is the second crack in classical physics: the precession of Mercury's orbit.

- When you do the calculations of what one body traveling in orbit around another should look like in general relativity, it turns out that the orbit precesses. This explains very nicely the anomaly of Mercury's 43 seconds of arc per century.

- Almost as soon as he published his theory of general relativity, Einstein demonstrated that orbital precession is one consequence and prediction of relativity, as confirmed by Mercury's anomalous orbit. As a result, Einstein dissolved the crack in classical physics.

- Today, orbital precession is used to explain the orbits of massive objects in very tight objects where the precession is easy to observe. Binary pulsars typically precess at several degrees per year—a fact that is very easy to measure astrophysically.

- Light travels in the straightest possible paths, but they, too, are not straight. Light follows the geometry of space-time, and it is bent as it travels near massive objects.

- The bending of light according to general relativity was predicted by Einstein and was verified by an eclipse that occurred in 1919.

- **Gravitational lensing**, the bending of light, has become a tool of astrophysicists for all kinds of processes because it can illuminate notions about the distant universe.

- In fact, gravitational lenses can act as giant cosmic telescopes to allow us to see objects much fainter and more distant than we would've been able to see without them.

- Another example is gravitational redshift, which is actually a kind of gravitational time dilation. The time of clocks runs slower when clocks are deep in a strong gravitational potential and a strong gravitational field.

- Today, collapsed stars, black holes, neutron stars, and white dwarfs all illustrate very clearly this phenomenon of gravitational redshift. In addition, the GPS satellite system demonstrates gravitational redshift.

- General relativity predicts ripples in space and time, called gravitational waves. They haven't been observed directly yet, but we're searching for them with giant interferometers.

- Another consequence of general relativity is black holes, which arise from the solution to Einstein's equation that measures what gravitation looks like around a spherically symmetric object.

- Finally, the large-scale structure of the universe has been revised to include our general relativistic studies of it. In addition, there's a possibility that dark energy could originate from Einstein's equations.

Important Terms

gravitational lensing: An effect caused by the general relativistic bending of light, whereby light from a distant astrophysical object is bent by an intervening massive object to produce multiple and/or distorted images.

tidal force: The force that is caused by differences in gravity from place to place; this force is what causes the tides—not gravity.

Suggested Reading

Wolfson, *EUP*, chap 33.9.

Questions to Consider

1. Why is Newtonian gravitation incompatible with special relativity?

2. Describe the principle of equivalence.

3. Name 2 new phenomena predicted by general relativity. Has either actually been observed?

Introducing the Quantum
Lecture 50

T he distribution of light from hot, glowing objects presented one of the early cracks in the worldview of classical physics. Max Planck proposed a resolution that implied that vibrations of atoms that produced light must be quantized—that is, they occur only with certain discrete energies. Similarly, Einstein explained the photoelectric effect by extending the quantization idea to light, which he said consisted of particle-like bundles of energy called photons. Experiments with the Compton effect showed that, in some circumstances, light interacts with electrons just as if it consisted of particles.

- The second of the new ideas of modern physics is an idea that grew not out of the mind of one person, as did relativity, but out of the minds and work of different approaches of quite a number of physicists through the early part of the 20th century.

- The work of these physicists culminated in a fairly firmly established theory of quantum physics by around the year 1930 and continued to grow throughout the remainder of the 20th century—and continues to grow today.

- The ultraviolet catastrophe was one of the cracks in classical physics. It was the prediction that there should be infinite energy emitted as any hot, glowing object approaches shorter and shorter wavelengths.

- In contradiction, the observation is that the energy peaks at a temperature-dependent wavelength.

- The resolution to the ultraviolet catastrophe came in 1900 from a German physicist by the name of Max Planck.

- Planck devised a formula that fit the observed distribution of radiation with wavelength and reduced to the classical prediction at long wavelengths—away from the ultraviolet catastrophe: $R(\lambda, T) = 2\pi hc^2/[\lambda^5(e^{hc/\lambda kT} - 1)]$, in which the h is called **Planck's constant** and is the fundamental aspect of quantum physics.

- If Planck's constant were zero, this formula would reduce to the classical prediction, and there would be no quantum effects at all. Planck's constant, therefore, tells us that our world is quantized.

- In SI units, Planck's constant is 6.63×10^{-34}, and its units are joules times seconds (J·s), which are also the units of something called action and of angular momentum.

- The mathematical implication of Planck's equation is that radiance is reduced to zero at small wavelengths. This is exactly what occurs in the observations, and it resolves the ultraviolet catastrophe.

- The physics implication is that the energy of the atomic and molecular vibrations—which are what cause a hot, glowing object to radiate energy called blackbody radiation—is **quantized**. That is, the energy exists in multiples of a certain fundamental amount, which is Planck's constant times the frequency of the vibrations.

- This physical implication solves the ultraviolet catastrophe because it means that the minimum amount of energy at very high frequencies—which correspond to very short wavelengths because frequency times wavelength is the speed of light—is so large that an atom with that amount of energy is very unlikely, so there is simply no radiation at very short wavelengths.

- The physical interpretation of Planck's resolution can be expressed in the equation $E = hf$, in which f is the frequency of an oscillatory process. In this case, the frequency of atomic and molecular vibrations is taking place in a hot, glowing object.

- E is the quantum of energy at that frequency, meaning that E is the minimum amount of vibrational energy these objects can have—a very nonclassical idea. Again, h is the proportionality constant, Planck's constant.

- In resolving the ultraviolet catastrophe, Planck introduced the idea of quantization. In this case, the vibrational energies of oscillatory processes like atomic and molecular vibrations are quantized.

- Another one of those cracks in classical physics involved the interaction of electrons and light, and the resolution of that is also going to involve quantization.

- This crack involves the interaction of light and electrons in which metal plates are in a sealed evacuated tube. When light is shined on those metal plates, electrons are ejected from the metal by the energy of the light, and electric current flows through this complete circuit.

- As argued in Lecture 44, it should take a while for the electrons to be ejected because it takes time for them to absorb enough energy from the light wave, where the energy is spread out.

- Increasing the light intensity should shorten that time because there is more energy available. The electron energy shouldn't depend on the light's wavelength.

- However, what actually happens is that the electrons are ejected immediately, so electron energy does depend on wavelength. In fact, electron energy increases with decreasing wavelength, and there is a cutoff wavelength at which electrons cannot be ejected at all.

- In 1905, Albert Einstein explained the cutoff frequency, or cutoff wavelength, below which frequency and above which wavelength no electrons are ejected. He said that light consists, not necessarily of waves, but of little bundles of energy called **photons**, or light particles.

- The formula for photon energy is the same formula that Planck created, $E = hf$, and it explains the photoelectric effect.

- There's a minimum energy needed to eject an electron from the metal because the metal is holding onto it with electrostatic forces. There is a minimum frequency required because the photon energy is directly proportional to minimum light frequency, which corresponds to the cutoff wavelength.

- The electron energy increases with increasing frequency because the photon energy, the energy of these light bundles that are coming in to eject the electrons, increases with increasing frequency.

- The current depends on the intensity, but the electron energy doesn't depend on the intensity because the current gives the number of photons.

- The electrons are ejected immediately because instead of this being an interaction where it takes a long time to absorb energy from a wave whose energy is all spread out, in fact, the interaction is with a little concentrated lump of energy that hits an electron and ejects it almost immediately—regardless of how intense or not intense the light is.

- The anatomy of this equation, $E = hf$, is a little different now that it is applied to light rather than to atomic vibration. In this case, f is a frequency of light, which is c/λ in terms of wavelength.

- In this case, E is not the energy of a molecular vibration but the energy of a photon, and h still represents Planck's constant.

- The reason quantum effects tend to be noticed with X-rays and not as much with radio waves is that at very low frequencies (radio waves), it takes so many photons to make up any appreciable beam with any appreciable power that you don't notice the discreteness of the energy bundled into these photons.

- As the frequency increases, however, it takes fewer and fewer photons to carry the same power, so you notice the quantization much more. Visible light is somewhere in between—sometimes quantum effects are seen, and sometimes they aren't.

- In the **Compton effect**, which was discovered by Arthur Compton in 1923, a photon collides with an electron, which recoils, and the photon bounces off. Because the electron gains energy while the photon loses energy, the photon has a lower frequency by $E = hf$ and, therefore, a longer wavelength. The Compton effect works in reverse, too.

- The Compton shift in wavelength is Planck's constant divided by the mass of the particle that's getting hit, in this case the electron, times the speed of light: h/mc.

- The Compton effect convinces us that photons are real and that light interacts in a particle-like way, which is more prominent at high frequencies.

Important Terms

Compton effect: An interaction between a photon and an electron in which the photon scatters off the electron and comes off with less energy. The effect provides a convincing demonstration of the quantization of light energy.

photon: The bosonic particle that mediates the electromagnetic force. An electromagnetic wave or field consists of a condensate of a large number of photons. Photons interact directly with any kind of particle that carries electric charge.

Planck's constant: A fundamental constant of nature, designated h (numerically equal to 6.63×10^{-34} J·s), that sets the basic scale of quantization. If h were zero, classical physics would be correct; h being nonzero is what necessitates quantum physics.

quantization: This word has a number of meanings in physics. One definition, the one most commonly used, refers to the fact that energy associated with any of the 4 fundamental forces comes in discrete packets—not in a continuum.

Suggested Reading

Rex and Wolfson, *ECP*, chap 23.1–23.3.

Wolfson, *EUP*, chap 34.1–34.3.

Questions to Consider

1. What's meant by the term "ultraviolet catastrophe"?

2. Explain the equation $E = hf$.

3. What is a photon, and how does the photon concept explain the photoelectric and Compton effects?

Atomic Quandaries
Lecture 51

I n 1912, Niels Bohr resolved the quandaries of the atom, proposing an atomic model in which electrons can move only in certain discrete orbits, corresponding to quantized levels of energy and angular momentum. This Bohr atom could emit electromagnetic waves only when electrons jumped from one orbit to another—thus solving both atomic quandaries. In 1923, Louis de Broglie showed why atomic energy levels should be quantized with an idea that gives matter, like light, both particle and wave aspects.

- Quantization has resolved 2 of the cracks in classical physics, and it also will take care of the following 2 problems with atoms:

  ○ First of all, atoms shouldn't exist because the electrons traveling around their nuclei are accelerated charges that should radiate electromagnetic waves. The energy for those waves should come at the expense of the electron's kinetic energy, and the electrons should quickly spiral into the nucleus, so atoms should not exist.

  ○ The other problem with atoms is that they emit only certain discrete wavelengths, and frequencies, of light instead of emitting a continuous spectrum of all colors.

- The idea of the atom may have originated with the ancient Greek philosopher Democritus, who thought that matter consisted of indivisible particles. In 1704, Newton echoed these thoughts.

- In 1897, British physicist Sir J. J. Thomson discovered the electron, the first clue that an atom is not indivisible, which led to the idea that there may be smaller parts that make up an atom.

- In the years around 1900, Marie and Pierre Curie discovered a bunch of subatomic particles while working with radioactive elements that emitted these particles. Atoms were no longer thought to be indivisible.

- In 1911, Ernest Rutherford discovered the atomic nucleus—that there is a tiny, massive center of an atom that contains nearly all the mass and the positive charge.

- At around the turn of the 20th century, the picture of the atom is a classical one that includes a center, the nucleus, and individual electrons traveling around it. The atom is mostly empty space. This picture of the atom is incorrect.

- The idea is that the electron travels around the nucleus in a circular orbit. Using terms from classical physics, the electron is bound by the electric force—the attractive force between the positive proton and the negative electron—and that any orbit around the nucleus should be possible.

- However, the observation is that spectral lines, the individual colors of light that an atom emits, have a patterned distribution.

- In 1884, Johann Balmer, a Swiss school teacher, devised an empirical formula that said if you look at the wavelengths for hydrogen, the simplest atom, they fit a simple pattern: The inverse of the wavelength, $1/\lambda$, is proportional to $1/4(1/2^2)$ minus $1/n^2$, where *n* is a number greater than 2:

$$\frac{1}{\lambda} = R_H \left(\frac{1}{2^2} - \frac{1}{n^2} \right).$$

- There is also a proportionality constant, R_H, called the Rydberg constant, which has the empirical value 1.1×10^7 m^{-1}. Its units are in inverse meters.

- The formula with $n = 2$ is called the Balmer series, and some of the other series have names as well. The Balmer series applies just to hydrogen; things become more complicated for other atoms.

- In 1913, Niels Bohr, the great Danish physicist, devised what's called the **Bohr atomic model**, which says that the electron is in a classical circular orbit around the proton. The electron doesn't radiate electromagnetic waves, and only certain orbits are allowed.

- Bohr's model states that the orbits that are allowed are those in which the angular momentum is quantized in integer multiples of Planck's constant divided by 2π. That quantity is given the name *h*-bar, and $\hbar = h/2\pi$.

- The model also says that the atoms are allowed to radiate electromagnetic waves or photons, but only when the electrons jump between allowed orbits.

- Therefore, when an electron is in a higher-energy orbit and jumps to a lower-energy orbit, it radiates electromagnetic energy. Then, it radiates the energy difference between the 2 orbits.

- The energy difference is radiated in the form of a single photon, and because energy is E, we can calculate the corresponding frequency by using the equation $E = hf$. We can then get the frequency f or the wavelength λ out of that.

- The Bohr model of the atom is $mvr = n\hbar$. Bohr also formulated an equation that allows us to find the energy levels of a hydrogen atom: $E = \dfrac{ke^2}{2a_0}\left(\dfrac{1}{n^2}\right)$.

- The Bohr atom has some ad hoc assumptions in it. It says that electrons are going around in circles, but for some reason they aren't radiating, unless they jump between allowed orbits. In addition, the allowed orbits are only certain discrete orbits where the angular momentum is quantized.

- The Bohr atom gives us a really good description of hydrogen, but it only works accurately for hydrogen and ions that have one other electron.

- In addition, it doesn't explain the fine structure in the hydrogen spectrum, details in which spectral lines are actually split into several lines.

In 1897, Sir J. J. Thomson discovered the electron, which was the first clue that the atom is not indivisible.

- The most important thing about the Bohr atom is that the assumption of quantized angular momentum simply isn't correct because that assumption is based on the idea that electrons are traveling in classical circular orbits, and it turns out that in most cases they aren't.

- In fact, some of the simplest hydrogen states don't have any angular momentum at all. Nevertheless, Bohr's model is somewhere between classical physics and the full quantum description of the atom, and his model explains a lot about the atom and quantization.

- Returning to our list of cracks in classical physics, special relativity, general relativity, and quantization solved 2 of the cracks.

- We have now also solved the problem of atoms by using the ideas of the quantization of angular momentum and Bohr's hypothesis that atoms don't radiate unless they jump between particular energy levels.

Lecture 51: Atomic Quandaries

- Therefore, we have finally resolved all of the cracks—2 of them by Einstein's relativity theories and 3 of them by quantization.

- In 1923, Louis de Broglie recognized, as Einstein had, that light has both wave (electromagnetic waves, per Maxwell) and particle (photons, per Einstein) aspects.

- What if we look at the momenergy for a massless particle? What if light consisted of massless particles? Then it would have the relationship that energy is equivalent to momentum times the speed of light.

- That would be the simple relationship for a massless particle, which is what a photon is. It's not a matter particle; it's a particle of light energy.

- If you put all of those things together, you find that the wavelength can be written as Planck's constant over the momentum for a photon: $\lambda = h/p$.

- De Broglie's hypothesis is if a particle with momentum p has associated with it a wave whose wavelength is given by the same formula, $\lambda = h/p$, what about a dual nature for matter? What about the possibility of matter waves?

- This is de Broglie's big philosophical idea, which takes solid, hard matter and makes it into something that has a wavelike nature. We don't know yet what this means for light to have both a wave nature and a particle nature.

- Physically, de Broglie's idea explains that the allowed orbits are those orbits that can accommodate an integer number of wavelengths. Therefore, Bohr posed an angular momentum quantization condition, which we used as a mathematical condition, but now we see physically why it works.

Important Term

Bohr atomic model: The atomic model proposed by Niels Bohr in 1913 in which electrons can only move in discrete orbits around the nucleus. When light is absorbed or emitted, the electron "jumps" from one orbit to another.

Suggested Reading

Rex and Wolfson, *ECP*, chap 24.1–24.2.

Wolfson, *EUP*, chap 34.4–34.5.

Questions to Consider

1. How does Bohr's atomic theory explain the origin of spectral lines?

2. How many quantized energy levels are available to the electron in a hydrogen atom? Do they span a finite or an infinite range of energies?

3. The ground-state energy in hydrogen is −13.6 electron volts. Why is there a negative sign?

Wave or Particle?

Lecture 52

A re light and matter particles or waves? Niels Bohr answered with his complementarity principle: Wave and particle descriptions are both needed for a full understanding of either light or matter, but they'll never be found in contradiction. Closely related is the Heisenberg uncertainty principle, which precludes measuring accurately and simultaneously certain properties of particles, such as position and momentum. Thus, quantization introduces a fundamental uncertainty, or indeterminacy, into what had been, in the Newtonian worldview, a completely deterministic physics.

- Is light a wave, or is it a particle? Similarly, is matter a wave, or is it a particle? Quantum physics puts us in this quandary, this wave-particle duality we somehow have to reconcile.

- It was in 1801 that Young first convinced the world that light consists of waves by showing that waves undergo interference. Earlier, Christiaan Huygens had thought light was a wave, but Newton, who had thought light was a particle, overruled by virtue of his scientific seniority.

- In 1905, Einstein explained the photoelectric effect by saying that light consists of photons, bundles of energy that behave, according to the Compton effect, like particles.

- Similarly, we have this wave-particle duality problem for matter. Democritus first hypothesized that matter consists of atoms, and then in 1911, Rutherford discovered the nucleus, proving that matter consists of atoms and subatomic particles.

- In 1923, de Broglie stated the possibility that matter has wave aspects—a notion that he uses to explain why Bohr's angular momentum quantization condition gives the quantized energy levels in the Bohr atom.

- If we were to do a 2-slit experiment, we would send a beam of incoming light—or, for that matter, electrons—into a barrier that has a couple of small slits in it.

- If light consisted of particles, we would expect the particles to go straight through the slits and form 2 bright spots on the screen opposite where the slits are—assuming the slits are wide enough that we are not going to see individual diffraction.

- On the other hand, if light consisted of waves, the waves would pass through the slits—which act as sources of circular wave fronts by Huygens's principle—and we would see circular wave fronts propagating outward until they hit the screen. You would see a pattern in which there were places where wave crests crossed and, in between, where wave troughs crossed.

- If you were to mark the lines where wave crests meet and where wave troughs meet, then you would get regions of particularly intense waviness, and you would see bright spots on the screen in a pattern of alternating light and dark bands that indicates that light is a wave.

- An idea that was summarized by Niels Bohr and described in a way that allows us to reconcile these seemingly conflicting points of view is called the principle of complementarity. Bohr said that the wave and particle aspects are complementary; there are 2 aspects of reality, and reality has both of them.

- In other words, if you do a wave experiment, you see wave behavior. If you try to set up an interference experiment to detect interference phenomena in electrons or in any other matter particles, then you are going to see wave behavior.

- If you do an experiment where you would expect to detect a particle—for example, Compton's scattering, where a photon scatters off an electron—then you will detect the particle aspect of that behavior.

Lecture 52: Wave or Particle?

- What you will never be able to do is have both aspects simultaneously, which would be a contradiction. You can't see both the wave and particle aspects of matter or light at the same time, and you choose what you're going to see.

- In a quantum system, anytime you try to detect particles, you're more likely to find a particle where the wave that is describing those quantum particles has a large amplitude and relatively few particles where the wave has a small amplitude and, in fact, zero particles where the wave has an amplitude of amplitude.

In 1911, Ernest Rutherford discovered the nucleus, which provided evidence that matter consists of atoms and subatomic particles.

- Because of this probabilistic relationship between waves, which is something we are able to calculate—the sort of randomness of where these particles are going to appear—nature is no longer certain and deterministic. Instead, it has become fundamentally probabilistic.

- Reality is fundamentally probabilistic, and this concept marks the end of Newtonian determinism. Because of this dual wave-particle nature of both matter and light, quantum mechanics and quantum physics become fundamentally probabilistic and, in some sense, indeterminate.

- A comparable relationship holds for the places where you would find electrons—or atoms, or protons, or whatever small particles you want to do an interference experiment with—and the corresponding wave amplitude.

- A wave is fundamentally more vague than a particle. You are able to describe where a particle is, and in classical physics, you also knew how the particle was moving, where it was headed, and how quickly it would get there.

- A wave, on the other hand, is kind of spread out. A wave has a vagueness to its position, and that vagueness translates into a vagueness in quantum physics. That vagueness manifests itself as a very famous principle in modern physics.

- Formulated by Werner Heisenberg in the 1920s, the uncertainty principle tells us quantitatively about this fundamental uncertainty in nature at the quantum level, and it's a fundamental uncertainty that arises ultimately from the fact that the nature of reality is dualistic.

- Heisenberg's quantum microscope is a famous thought experiment in which he observes an electron. His microscope consists of a light source that sends out a single photon, which bounces off the electron and into a detector.

- In this thought experiment, we use things like where the detector is positioned and how much energy is involved to measure the properties of the electron. We try to figure out where it was and how fast it was moving.

- In these types of situations, there exists a tradeoff between trying to measure position accurately, which you could do at the expense of losing information about motion velocity or, in the context of the uncertainty principle, about momentum.

- The uncertainty principle says if you multiply the uncertainty in position and the uncertainty in momentum, it has to be bigger than some certain value, which is, basically, Planck's constant: $\Delta x \Delta p \geq \hbar$.

- If you make Δp very small, that means you know the momentum accurately, and Δx has to be correspondingly big. If you make Δp zero, Δx approaches infinity. Similarly, if you make Δx very small, you know the position very accurately, and Δp becomes very large so that you don't know the momentum at all.

- The uncertainty principle tells us there are these properties—momentum and position, energy and time—that we can't measure simultaneously.

- Therefore, particles don't have well-defined values of position and momentum, so Newtonian determinism, classical Newtonian trajectories, and even the circular orbits that Bohr wanted electrons to have in the atom are all dispelled in exchange for the ideas of quantum mechanics, which is much more complex.

- The standard understanding of quantum physics held by most physicists, but not all philosophers of science, is called the **Copenhagen interpretation**, which says that it makes no sense to talk about things you can't measure.

Important Term

Copenhagen interpretation: The standard view of the meaning of quantum physics, which states that it makes no sense to talk about quantities—such as the precise velocity and position of a particle—that cannot, even in principle, be measured simultaneously.

Suggested Reading

Rex and Wolfson, *ECP*, chap 23.4.

Wolfson, *EUP*, chap 34.6–34.7.

Questions to Consider

1. An electron passes through a 2-slit system. Can you tell which slit the electron went through from where it lands on the screen?

2. Is the uncertainty principle a limitation on our ability to measure properties of the physical world, or does it reflect a deeper indeterminacy?

3. Describe the central tenet of the Copenhagen interpretation of quantum physics.

Quantum Mechanics
Lecture 53

E rwin Schrödinger developed an equation that describes the behavior of matter in terms of associated waves. The Schrödinger equation yields a wave function that completely characterizes the state of a given particle. The wave function gives probabilities for finding the particle in a particular location or moving with a particular speed. The Schrödinger equation leads to energy quantization anytime forces confine a particle to a limited region. Making quantum mechanics consistent with the principle of special relativity leads to startling new phenomena such as antimatter and the spin of elementary particles.

- Given that there is a wave-particle duality applied both to light and matter, if you can calculate what the wave looks like or figure out what the wave is going to do, then based on this probabilistic interpretation, you can find out what the particles of the wave are going to do and where to find them.

- The **Schrödinger equation** was developed in the 1920s by Erwin Schrödinger, and it is the equation that describes the waves associated with matter. It's analogous to Maxwell's equations for electromagnetic waves.

- The Schrödinger equation gives us a wave function, which is given the symbol ψ, the Greek letter psi, for a given particle. In 1 dimension, it's a function of position x.

- We're going to address only the time-independent Schrödinger equation, and we're going to address only how the wave function depends on position and not on time—although it does.

- It's not quite the amplitude of the wave function that gives the probability of finding a particle at some position; it's the square of the amplitude. Although the amplitude of the wave can be positive or negative, its square is always positive.

- The square of the wave function, which is a function of position x, gives the probability of finding a particle at that position. We're going to express this in terms of mathematical operators acting on ψ.

- The position operator x is proportional to the wave function ψ, so you can multiply x by ψ. The momentum operator p is proportional to the rate of change of the wave function with respect to x, and the symbol d/dx is an operation that signifies the calculus limit of Δ, the rate of change. The kinetic energy is $1/2mv^2$, which can be solved to become $p^2/2m$. Remember also that $\hbar = h/2\pi$.

- Using substitution, the kinetic energy is $-\dfrac{\hbar}{2m}\dfrac{d^2\psi(x)}{dx^2}$.

- The potential energy turns out to be $U(x)\psi(x)$.

- Combining the kinetic energy and the potential energy, we get the time-independent Schrödinger equation:

$$-\frac{\hbar}{2m}\frac{d^2\psi(x)}{dx^2} + U(x)\psi(x) = E\psi(x).$$

- The Schrödinger equation is basically a statement of conservation of energy. It says that kinetic energy plus potential energy equals the total energy—except instead of the energies, the energy operators are acting on the wave function, which is equal to the total energy acting on (multiplying) the wave function.

- The special case for the Schrödinger equation is that of a completely free particle—one that is subject to no forces and that should obey Newton's first law. In the classical sense, it should move at a constant speed with no change in its motion—no change in potential energy with position.

- In the case of a free particle, the Schrödinger equation becomes a simpler equation that simply has the kinetic energy operator equal to the total energy operator acting on the wave function.

- Newton's second law and the Schrödinger equation for a free particle are 2 mathematical statements that are analogous.

- The solution we found in Lecture 17 was that the position x of a mass on a spring was a sinusoidal function of time. It could have any arbitrary amplitude, and it could either be a cosine or a sine of some angular frequency times time. The angular frequency turned out to be the square root of k, the spring constant, divided by m, the mass.

- The solution for the wave function is analogous to the solution from classical physics: $\psi = A \frac{\sin}{\cos}\left(\sqrt{2mE}\,x/h\right)$.

- The Schrödinger equation gives the solution for a particle's wave function ψ, which gives the probability of finding a particle in a given wave. That probability is proportional to the square of that wave function.

- The Schrödinger equation is based on nonrelativistic physics, so it's valid only for particles moving much slower than light.

- For confined particles—like a particle in a box—we find that quantization arises naturally by the requirement that those waves fit within the box.

- With a particle confined between rigid walls, only certain waves fit between the walls. Each wave corresponds to a different energy, thus yielding energy quantization and, at the same time, showing that there's a nonzero minimum energy.

- The **correspondence principle** is another principle that was formulated by Bohr and says that quantum mechanics agrees with classical physics, but only in the limit of very large quantum numbers.

- An interesting thing that occurs in quantum physics—which is not truly classical physics because waves can superpose—is that 2 waves can exist in the same place. We can have 2 waves that, in fact, add together to give a composite wave.

- A quantum harmonic oscillator has a potential energy function that is quadratic in position, and we can write that as $1/2 m\omega^2 x^2$, where ω is the angular frequency of that system.

- It's possible to solve mathematically the Schrödinger equation for that state, and what you get are, in fact, evenly spaced energy levels starting at $1/2 hf$ or $1/2 h\omega$ because they're all transitions of magnitude hf.

- **Quantum tunneling** represents the nonzero probability of finding a particle where classical energy conservation prohibits it.

- When confining walls aren't perfectly rigid—as in the quantum harmonic oscillator—the wave leaks into the region beyond the walls and results in a nonzero probability of a particle being found, where classical physics says it's energetically impossible. This is quantum tunneling, and it plays an important role in many natural and technological systems.

- Quantum physics particles also live in 2 and 3 dimensions. In 2 dimensions, you can fit wave functions in both dimensions, so you have 2 quantum numbers that describe how many waves there are.

- Quantum mechanics and general relatively have not been reconciled, but quantum mechanics and special relativity were reconciled decades ago by Richard Feynman and others, who noticed that the Schrödinger equation was based on $K = 1/2mv^2$, the formula for Newtonian kinetic energy.

- In 1928, P. A. M. Dirac theoretically discovered the existence of antiparticles, or antimatter. He then developed a relativistic quantum theory of the electron with what's called the Dirac equation.

- Another idea that stems from relativity is that particles have to have an intrinsic angular momentum, called **spin**. Most of the particles that make up matter are called fermions, and their spin is a 1/2-integer multiple of $\hbar$.

- These particles, called spin-1/2 particles, obey something called the **Pauli exclusion principle**, which states that 2 particles can't be in the same quantum state.

- Then there are bosons, particles with integer spin, and they can be in the same state. Photons have integer spin.

- All of these notions follow from making quantum mechanics consistent with the principle of relatively. Quantum mechanics and its reconciliation with relativity give rise to new phenomena that are important in understanding, particularly, atoms.

Important Terms

correspondence principle: A principle formulated by Niels Bohr that says quantum mechanics agrees with classical physics, but only in the limit of very large quantum numbers.

Pauli exclusion principle: The impossibility of putting 2 fermions into the same state. It is this property of fermions that makes them matter particles—they take up space. Bosons, in contrast, do not obey the exclusion principle and can be squeezed together without limit.

quantum tunneling: The surprising phenomenon by which a quantum particle can sometimes pass through a potential energy barrier that would (under classical physics) ordinarily be expected to block it.

Schrödinger equation: The equation discovered by Erwin Schrödinger that controls how the quantum wave function behaves over time.

spin: The intrinsic angular momentum of an elementary particle, which is the property of appearing to have attributes of tiny spinning balls, even though they possess no size at all. The rates of spin of elementary particles are measured in terms of a quantity called h-bar ($\hbar$) and come in any integer or 1/2-integer multiplied by this rate.

Suggested Reading

Wolfson, *EUP*, chap 35.

Questions to Consider

1. In Schrödinger's theory, a particle with precisely known momentum is described by an infinitely long sinusoidal wave. Can you say anything about where such a particle is located? How is your answer consistent with the uncertainty principle?

2. A particle confined to a rigid box is in its ground state. Where are you most likely to find the particle? How is your answer at odds with classical physics?

3. What is quantum tunneling?

Atoms

Lecture 54

The electric force between the proton and electron in a hydrogen atom acts like confining walls of a box, resulting in quantized atomic energy levels. Thus, the Schrödinger equation explains the energy quantization proposed in Bohr's atomic model. The Pauli exclusion principle states that no 2 electrons can be in the same quantum state. Because there are 2 possible states for an electron's spin, at most 2 electrons are in otherwise identical quantum states. The exclusion principle also dictates the arrangement of electrons in more complex atoms. Quantum mechanics is the basis of chemistry.

- In the Schrödinger equation, we had the kinetic energy operator, the potential energy operator, and the total energy. The different potential energies determine the way a system behaves.

- In a 1-dimensional box (a particle in a box) and in a harmonic oscillator, the mathematics of the Schrödinger equation results in discrete quantized energy levels—different energy levels that are available to the system.

- You can think of an atom as another system that consists of 2 point charges—the proton, $+e$, and the electron, $-e$—for the simplest atom, hydrogen.

- The potential energy of 2 point charges is kq_1q_2/r. In this case, q_1 is $+e$ and q_2 is $-e$, so we get $-(ke^2)/r$ for the potential energy function that we can substitute into the Schrödinger equation.

- A diagram that plots energy as a function of position shows a confining potential, just like the harmonic oscillator has a confining potential. You shouldn't be surprised that we also get quantized energy levels in the case of the atom.

- Bohr's answer, although it produced the right energy levels, isn't quite right. Let's look at the potential energy function for the hydrogen atom in particular.

- Assume there's a proton fixed at $r = 0$, radial position zero, and the electron is bound by electric force. We're also going to add the Coulomb potential $-\dfrac{ke^2}{r}$ to the Schrödinger equation.

- Mathematically, you need to use spherical coordinates—x, y, and z would be very cumbersome to work with—because the hydrogen atom has a spherical symmetry. The Coulomb potential, the electrostatic potential, has a radial symmetry about it from the point charge.

- **Spherical coordinates** are used to describe the position of a point in terms of its radial distance from some origin and then 2 angles θ, which describe its orientation relative to some axis, and φ, which describes its orientation around that axis.

- There are 3 numbers to describe position in 3-dimensional space, except instead of being x, y, and z, they are r, θ, and φ—and they have rather different mathematical manifestations as a result.

- The Schrödinger equation in spherical coordinates for the hydrogen atom is:

$$-\frac{\hbar^2}{2mr^2}\left[\frac{\partial}{\partial r}\left(r^2\frac{\partial \psi}{\partial r}\right) + \frac{1}{\sin\theta}\frac{\partial}{\partial \theta}\left(\sin\theta\frac{\partial \psi}{\partial \theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2\psi}{\partial \phi^2}\right] -$$

$$\frac{ke^2}{r}\psi = E\psi.$$

- The first 3 terms comprise the kinetic energy, which is what we had with the Schrödinger equation—except it now looks much more complicated because we're using a 3-dimensional coordinate system in which the different coordinates behave differently.

- When a hydrogen atom is in spherically symmetric states, it means that nothing depends on the angle, so all the terms involving θ (the angle) and φ (the longitudinal angle) disappear. They only depend, therefore, on r.

- As a result, the Schrödinger equation becomes simpler because there's no dependence on angular position: $\psi = Ae^{-r/a_0}$, in which A and a_0 are undetermined constants.

- The hydrogen ground state is the lowest possible energy for the hydrogen atom and is given by the equation $a_0 = \dfrac{\hbar^2}{mke^2}$, in which a_0 is the Bohr radius.

- The Schrödinger equation agrees with Bohr's theory, and it also happens to agree with observations.

- Of all the points in space, where is the electron in a hydrogen atom in the ground state most likely to be found? Quantum physics tells us that the answer is where the ψ function is biggest, and that's at $r = 0$, which is right inside the nucleus—in the center of the proton.

- More importantly, at what radial distance from the proton are you most likely to find the electron? The radial probability distribution shows that the place you're most likely to find the electron in the hydrogen ground state is 1 Bohr radius from the nucleus.

- There are other states the hydrogen atom can be in that are similar to the ground state, except they have higher energies. All these states share something important with the ground state: Because they have no angular dependence, they have no angular momentum either.

- You can determine from these various states that Bohr's theory is completely wrong on one point: There is no angular momentum in these states, and yet he developed his theory on the idea of quantized angular momentum.

- On the other hand, electrons can have angular momentum, and the energy it would take to take the electron entirely away from a hydrogen atom is given by $E_n = -\frac{1}{n^2} \frac{\hbar^2}{2ma_0^{\;2}} = \frac{E_1}{n^2} = \frac{-13.6\text{eV}}{n^2}$. This is called the principle quantum number, and it describes the energy of the electron.

- For a given *n*, there are *n* possible values for what's called the electron's "orbital" angular moment. If it were a classical particle, that value would be the angular momentum associated with its orbit.

- These states are distinguished by the quantum number *l* (the actual angular momentum is *L*), which range in value from 0 to *n*.

- You can calculate the actual angular momentum with the formula $L = \sqrt{l(l+1)}\hbar$, which is how you quantize angular momentum.

- If *l* is zero—and that's one of the allowed states—it results in those spherically symmetric *s* states that have no angular momentum.

- So far, then, we have 2 quantum numbers (*n* and *l*), but we're going to need 3 because we're in 3 dimensions.

- Another quantization that occurs is called **space quantization**, which results from the fact that angular momentum is a vector and that the direction of the angular momentum is also quantized—as described by the magnetic quantum number m_1: $L = m_1\hbar$.

- These 3 quantum numbers—the principle quantum number n, the orbital quantum number l, and the magnetic quantum number m_1—describe the state of the electron in hydrogen almost fully.

- There is one other quantum number that comes from that idea that when you try to couple quantum mechanics with relativity, electrons have an intrinsic angular momentum called 1/2-integer spin because the projection of that angular momentum on a given axis is $\pm 1/2 \hbar$.

- Therefore, there are 2 possible spin states. In addition to all those other states specified by the 3 quantum numbers, a given electron can have spin up or spin down, which is its spin relative to some particular axis. The quantization of spin angular momentum is given by: $S = \sqrt{s(s+1)}\hbar.$

- You can picture an atom as kind of an electron probably having electron probability cloud around it. In this image, there is an electron going around a proton, but as long as it has some angular momentum, the proton moves relative to it in a way that makes it look like there's a magnetic field from the moving charged proton.

- The magnetic fields originating from the interaction of the electron's orbital motion and spin result in slightly altered energy levels, which give rise to complicated structures in atomic spectra that simply could not be explained using something even as sophisticated as Bohr's atomic theory. You can apply this to multi-electron atoms by figuring out their electron configurations.

- All of chemistry is determined ultimately by the Pauli exclusion principle—that no 2 fermions (including electrons), particles with 1/2-integer spin, can be in the same state—and the way you can fit electrons into different levels with different quantum numbers, which leads to the periodic table of the elements.

Important Terms

space quantization: A quantization that results from the fact that angular momentum is a vector and that the direction of angular momentum is quantized—as described by the magnetic quantum number m_l: $L = m_l \hbar$.

spherical coordinates: Used to describe the position of a point in terms of its radial distance from some origin and then 2 angles θ, which describe its orientation relative to some axis, and φ, which describes its orientation around that axis.

Suggested Reading

Rex and Wolfson, *ECP*, chap 24.3–24.5.

Wolfson, *EUP*, chap 36.

Questions to Consider

1. Energy levels for a particle confined to a box are quantized because of the particle's confinement. Energy levels for an electron in an atom are also quantized. What provides the confinement in the atom?

2. Comment on this statement: "The electron in the ground state of hydrogen is located 1 Bohr radius from the nucleus."

3. How does the exclusion principle account for the existence of different chemical elements?

Molecules and Solids

Lecture 55

A toms join to make molecules, using a variety of physical mechanisms that includes a direct electrical attraction (ionic bonding) and a quantum-mechanical sharing of electrons (covalent bonding). Molecular energy levels, like those in atoms, are quantized. When many atoms join to make solids, the quantized energy levels get so close that they become essentially continuous bands, separated by energy gaps. Exploiting this band structure allows us to engineer the semiconductors at the heart of all our electronic devices.

- There are several ways molecules can bond together.

  ○ In an ionic bond, one atom that is prone to giving up an electron and another atom that's prone to receiving an electron come together. The one atom becomes energetically preferable for the one atom to lose its electron to the other, resulting in 2 charged particles. Ionic bonds form bonds that are very tightly held together. Ordinary salt, NaCl, is ionically bonded.

  ○ A covalent bond is a bond in which atoms share electrons among themselves, and it's the sharing of the electrons that gives rise to a lower-energy state that represents the binding together. The electrons pair with opposite spins, and they're in a single molecular orbital state, but because they have opposite spins, they can still obey the exclusion principle. Water is covalently bonded.

  ○ There are hydrogen bonds that occur when a tiny nucleus of hydrogen, which is just a single proton, is able to nestle down close to other structures. Hydrogen bonds are fairly weak; they bond water molecules to form ice. They also bond the DNA helix into its double-helix structure.

- ○ A van der Waals bond is a very weak bond associated with the dipole-dipole interaction.

- ○ Finally, there's the metallic bond, which is the bond that holds metal atoms together in a large chunk of metal. It's a bond that is forged by free electrons, which are free to roam throughout the metal and exert a bonding effect that holds the metal together.

- Molecules can rotate, and because of this, they have angular momentum. The angular momentum is $I\omega$, and it turns out to be quantized by a similar formula that was introduced for quantization of angular momentum, $L = \sqrt{l(l+1)}\hbar$.

- Molecular rotation is quantized because the rotational energy is the analog of $1/2mv^2$. It's $1/2I\omega^2$, and if you work that out using the quantized angular momentum formula, you get $E_{\text{rot}} = \dfrac{\hbar^2}{2I}l(l+1)$ for $l = 0, 1, 2, 3, \ldots$.

- There's also the possibility of molecular vibration: $E_{\text{vib}} = (n + 1/2)\hbar\omega$.

- Similarly, the simple harmonic oscillator in quantum physics has energy levels given by $(n + 1/2)hf$ of the vibration or $(n + 1/2)h\omega$.

- The lowest energy state has $1/2hf$ of energy, and they increase in steps by hf. The levels are separated $\Delta E = hf$. That tells you that transitions among vibrational energy levels in an atom are going to involve photons of energy hf.

- The spectrum of a molecule is somewhat complicated, and there are a whole host of transitions possible between energy states in rotating and vibrating molecules, which lead to evenly spaced spectral lines because of the simple harmonic nature of them.

- The way the different levels use different transitions to go from different levels occurs according to the rules $\Delta l = \pm 1$ and $\Delta n = \pm 1$.

- Molecules are a little heavier than atoms, and their vibrational frequencies are therefore a little bit lower in their rotational frequencies, so they turn out to involve mostly infrared photons. Molecular spectra tend to be in the infrared region of the spectrum.

- Condensed-matter physics is an area of physics that describes solid-state physics by what's called band structure.

- When we bring atoms together, we start with 2 atoms that are initially completely separated. When the atoms were far apart, 2 electrons could be in the $1s$ state and still be completely independent atoms.

- But as the atoms come together and those electrons sort of sense each other and begin to be shared, they can't both be in exactly the same state.

- Therefore, the energy level has to split, and they have to become 2 different energy levels as the atoms move closer together. If we bring more atoms together—5, for example—what has to happen is each of those energy levels splits into 5.

- Solids, as opposed to molecules, are not made of a few atoms—they're made of zillions of atoms. When we bring that many atoms close together, we get a gigantic number of levels of splitting, which creates basically a band of available energy states that are available to the electrons in that solid.

- Between 2 bands is an energy gap. No electron is allowed to have an energy in that gap for the same reason that no electron was allowed to have an energy between the allowed orbits in the Bohr model or the allowed orbits that are allowed by the solutions to the Schrödinger equation for that particular atom.

- If a material is an electrical insulator, it's a material in which the states in the lower of the 2 outermost bands are completely occupied by electrons. The other state is completely unoccupied.

- We call the lower one the valence band because it's the valence electrons that participate in bonding with other atoms, and the upper one is called the conduction band.

- **Electrical insulators** have completely occupied bands—and an energy gap before there are any unoccupied states—so it takes a large amount of energy to promote electrons into the unoccupied states so that they can conduct.

- In electrical conductors, it takes very little energy to promote an electron to a new unoccupied state, so conductivity comes very easily.

- Semiconductors are materials that are perfect insulators at absolute zero, at which point they have a fully occupied lower band, a fully unoccupied upper band, and an energy gap. Semiconductors are like insulators—not conductors.

- However, if the temperature is greater than absolute zero, the gap is so small that random thermal vibrations can promote occasional electrons into the conduction band.

- Intrinsically, semiconductors are not great conductors, but they do conduct electricity. More importantly, semiconductors are at the heart of modern electronics because their electrical properties can be manipulated by manipulating their band structure.

- Good conductors are created in 2 different ways: by adding extra electrons, which increase the quantity of free electrons, to form what is called an *N*-type semiconductor; or by adding a material, called a *P*-type semiconductor, to increase the number of free positive charges in a structure.

- Basically, a positively charged hole develops, a place where an electron is missing, and a nearby electron that is still bound in that structure can fall into that hole, which has effectively moved and caused conduction.

- Combining pieces of *N*- and *P*-type materials together creates a *P-N* junction, which acts like a one-way valve. It conducts electricity in one direction, but not the other—a concept that is at the heart of almost all modern electronic devices.

- Superconductors are solids that, below a certain temperature, conduct electricity with absolutely no resistance.

- In a superconductor, 2 widely separated electrons are quantum-mechanically paired so that they share a common wave function and carry electric current with no loss of energy.

- Another remarkable solid is **degenerate matter**, which is matter that is so tightly crammed together that basically all the particles act as one.

Important Terms

degenerate matter: Matter that is so tightly crammed together that basically all the particles act as one.

electrical insulator: An insulator that has completely occupied bands—and an energy gap before there are any unoccupied states—so it takes a large amount of energy to promote electrons into the unoccupied states so that they can conduct.

Suggested Reading

Wolfson, *EUP*, chap 37.

Questions to Consider

1. What accounts for the high melting point of sodium chloride (ordinary salt)?

2. As you bring atoms close together, what happens to the energy levels of the individual atoms? How does this account for the energy bands in solids?

3. How does band theory account for the difference between insulators, conductors, and semiconductors?

Lecture 55: Molecules and Solids

The Atomic Nucleus
Lecture 56

At the heart of the atom is the tiny but massive nucleus. Stable nuclei require a combination of protons and neutrons. Different chemical elements have different numbers of protons in their nuclei, but nuclei of the same element can have different numbers of neutrons, resulting in different isotopes of the same element. Many combinations of neutrons and protons aren't stable; such nuclei are radioactive and decay with characteristic half-lives. Nuclear radiation is hazardous to living things but also has beneficial uses.

- Perhaps more than any other area of physics, nuclear physics is deeply entwined with history, international politics—and, indeed, with the fate of the world.

- The discovery of the nucleus began in about 1896 with Henri Becquerel's discovery, quite inadvertently, of the phenomenon of radioactivity.

- J. J. Thompson in 1897 discovered the electron, recognizing that atoms are not indivisible.

- Out of Thompson's discovery came the plum-pudding model of the atom, in which the atom consists of a pudding of positive charge with individual electrons within it.

- Marie and Pierre Curie discovered the radioactive elements radium and polonium. The high-energy particles coming out of these radioactive materials gave us tools to probe the nature of the nucleus.

- Ernest Rutherford, Ernest Marsden, and Hans Geiger did a series of experiments from 1909 to 1911 in which they bombarded a thin foil with some of these high-energy particles from radioactive radium.

- By seeing how the particles recoiled, they were able to conclude that the atom consisted of a very tiny, massive center—the nucleus—surrounded by mostly empty space in which the electrons were orbiting.

- From that came the classic solar-system model of the atom, and you've learned in studying quantum physics that this model is not quite correct, but it certainly still is correct to think of the nucleus as a tiny, but very massive, particle at the center of the atom.

- The last step in this discovery is Sir James Chadwick's discovery of the neutron in 1932, understanding that the nucleus consists of protons and neutrons.

- The nucleus is positive, and it attracts electrons. In a neutral atom, the number of electrons equals the number of protons.

- Chemical properties are determined by the number of electrons, and the number of electrons is determined by the number of protons, so the number of protons in the nucleus determines the chemical properties of a particular atom. We call the number of protons in an atom the **atomic number** and give it the symbol Z.

- All atoms with the same atomic number belong to the same chemical element. They have the same number of protons in their nucleus and, therefore, the same number of electrons surrounding them and the same chemistry.

- However, the number of neutrons in a nucleus is not determined by anything chemical. You can add or take away a neutron and still have the same number of protons, the same number of electrons, and the same chemistry.

- Therefore, it is possible to have atoms of the same element that have different numbers of neutrons; those are called different **isotopes** of the same element.

- Isotopes have the same atomic number *Z*, so they're chemically similar, but they're not absolutely identical because they have slightly different masses, and that makes them behave slightly differently.

- Isotopes have different mass numbers, however. The **atomic mass**, which we designate by *A*, is the total number of nucleons. A **nucleon** means either a proton or a neutron. The mass number is approximately the mass of the nucleus.

- Helium-4, the most common kind of helium, has 2 protons and 2 neutrons. Helium-3 is still helium because it has 2 protons, but it only has 1 neutron.

- Helium-4 and Helium-3 behave similarly chemically, but they behave quite differently physically because of their different masses—and because one is a boson and one is a fermion.

- There are 2 forces involved in building a nucleus: an electric force and a nuclear force.

- The electric force acts repulsively between protons and tends to tear the nucleus apart. It is weak, but it extends over fairly long distances, as we know from our study of electricity. It falls off like $1/r^2$.

- Then there's a nuclear force that acts attractively between all the nucleons. It's actually a residual effect of the strong force that binds the quarks together to make protons and neutrons.

- The nuclear force is very strong, but it's also in a very short range. It falls off approximately exponentially, and the consequence of this is that larger nuclei need more neutrons than they have protons.

- What happens to nuclei that are not stable? Eventually, they decay; that is, they spew out some particles and turn into something else.

- One mode of decay that is favored by proton-rich nuclei—which are the ones that have too many protons—is to emit what's called an alpha particle in a process known as alpha decay.

- An alpha particle is just another name for a helium-4 nucleus, which is a version of helium with 4 total nucleons and a mass number of 4.

- Beta decay takes place in nuclei that typically have too many neutrons. The neutron decays to a proton, to an electron, and to an elusive particle that is chargeless and almost massless called a **neutrino**.

- The process is called beta decay because the high-energy electron that results is called a beta particle.

- Gamma decay occurs when a nucleus gets excited—analogous to the way an atom can get excited by its electrons rising to higher energy levels.

- In gamma decay, an excited nucleus drops to a lower energy level, emitting a high-energy photon called a gamma ray. The nucleus itself doesn't change; all it does is shed some energy.

- Where do the radioactive isotopes come from? Some of them are made by nuclear reactions in particle accelerators and nuclear reactors, but others exist in nature.

- How rapidly do radioactive materials decay? There is a certain time in which roughly half the nuclei in a given sample will decay, which is called the half-life.

- Half-lives range from fractions of a second to billions of years, and the number of nuclei after a given time is described by the number of nuclei you started with (N_0) times 2 raised to the power $-t$ over the half-life: $N = N_0 2^{-t/t_{1/2}}$.

- **Radiocarbon dating** is the process of using radioisotopes to date ancient objects. In particular, archeologists use carbon-14, which is formed in the atmosphere by cosmic rays, taken up by plants and then by animals.

- Uranium and potassium isotopes can be used to date objects back billions of years. Geologists use these isotopes to date rocks.

- Activation analysis makes ordinary materials briefly radioactive, and by detecting the signature of the radiation, you can tell what kinds of materials exist in an object.

- The SI unit of radioactivity is called the becquerel (Bq), which is named after the man who discovered radioactivity. A becquerel is simply a unit of the rate at which radioactive material is decaying.

- A gray is a unit of radiation dose. One gray means a material exposed to radiation has absorbed 1 J of energy from the radiation.

- The sievert (Sv) is the measure of biological radiation dose. It means the same thing as the gray; 1 Sv is 1 J/kg.

Important Terms

atomic mass: The sum of the number of nucleons (protons plus neutrons) in the nucleus of an atom (or of all the nucleons in a molecule). The atomic mass number is written to the upper left of the chemical symbol for the element. In all nuclear transformations, the atomic mass is conserved.

atomic number: The total number of protons in an atom's nucleus and, hence, the number of electrons in a neutral atom. Determines what element an atom belongs to.

isotope: Atoms with identical numbers of protons and electrons that differ only in their number of neutrons. Since the number of electrons determines the atom's chemical behavior, all isotopes of an element behave identically in forming molecules with other atoms; the sole difference is their mass.

neutrino: The lightest of the subatomic particles with masses on the order of 1 millionth of the electron. Neutrinos only interact via the weak nuclear force and, as such, pass through matter with ease. They accompany all beta decays and can be used to peer directly at the nuclear furnace at the core of the Sun.

nucleon: A generic name for neutrons and protons, the constituents of nuclei.

radiocarbon dating: The use of the radioactive isotope of carbon, carbon-14, to determine the age of an object. Carbon-14 is produced in the atmosphere when cosmic rays strike nitrogen atoms in the air.

Suggested Reading

Rex and Wolfson, *ECP*, chap 25.1–25.4.

Wolfson, *EUP*, chap 38.1–38.3.

Questions to Consider

1. Why are isotopes of the same element chemically similar?

2. Why do heavier nuclei have more neutrons than protons?

3. By roughly what factor does a radioactive material's activity decline in 10 half-lives?

Energy from the Nucleus
Lecture 57

The interplay between nuclear and electric forces results in a variation of nuclear binding energy among nuclei. Energy can be released by breaking up heavier nuclei—the process of nuclear fission—or joining lighter nuclei, called nuclear fusion. Both types of nuclear reactions release millions of times more energy than chemical reactions. Fission is the energy source for nuclear power plants and nuclear weapons. Fusion is the process that powers the Sun and other stars and is responsible for the creation of all but the lightest chemical elements.

- In an atomic nucleus, there are competing forces that both hold the nucleus together and tend to drive it apart. There's an electric force acting repulsively between all the protons in the nucleus. A quantity called binding energy measures how tightly bound a nucleus is.

- In the nucleus, protons repel each other, which tend to make the nucleus fly apart, and that is decreasing the binding energy—making it less tightly bound. The electric force is weak, but it extends over long distances.

- There's also an attractive nuclear force acting in between all the nucleons. It increases the binding energy because it's holding the nucleus together. It's strong, but its range is short.

- One conclusion is that more massive nuclei tend to have more neutrons than protons, whereas lighter nuclei tend to have equal numbers of them.

- The middle-weight nuclei are the most tightly bound because if you start adding more protons beyond the middle-weight nuclei, you begin to lower the binding energy—the strength of the binding—because you've got more of that repulsive force.

- If you start out with a very small nucleus and start adding nuclei, you get nucleons, which increase the binding energy. It's only when the nucleus starts getting larger that the repulsive effect of the protons lowers the binding energy.

- There is a graph called the **curve of binding energy** that plots nuclei by weight on the horizontal axis, starting with hydrogen, the lightest, to uranium, the heaviest naturally occurring nucleus. On the vertical axis is the binding energy per nucleon.

- There are 2 ways that you can combine or alter nuclei in order to release energy.

  ○ One is to fuse very light nuclei in a process known as **fusion**. If you take light nuclei, which are not very tightly bound because there isn't much nuclear force (because there aren't many nucleons), and you bind them together into a somewhat heavier nucleus, energy will be released as that heavier nucleus is more tightly bound.

  ○ The other way to release energy is to break heavy nuclei apart in a process known as **fission**. Nuclear fission takes heavy nuclei—because they aren't quite as tightly bound as middle-weight nuclei—and breaks them apart into 2 middle-weight nuclei. There's tighter binding with the 2 middle-weight nuclei, so you have released some energy in the process.

- In principle, you can do these energy-releasing reactions with any nuclei. In practice, however, it's difficult. In particular, fission is very difficult to make happen under most circumstances with just any heavy nucleus.

- Nuclear fission was discovered in the late 1930s. Throughout the 1930s, a number of physics labs around the world were doing experiments in which they were bombarding nuclei with neutrons in order to study the nuclei.

- A lab in Germany with chemists Otto Hahn and Fritz Strassmann bombarded uranium with neutrons and discovered that their final product was contaminated with middle-weight nuclei—particularly barium, which has atomic number 56.

- Physicist Lisa Meitner and her nephew Otto Frisch calculated the energy that would be released from the fact that uranium seemed to be splitting into 2 through these bombardments.

- They looked at the electrostatic repulsion of the 2 nuclei that would be created and at the Einstein $E = mc^2$ mass difference, and they came up with a colossally large number, in both cases, for the energy that would be released in this process.

- By 1942, a team led by Enrico Fermi in Chicago had built the first successful nuclear reactor. In 1945, one nuclear weapons test was followed by 2 nuclear weapons being used against Japanese cities.

By 1942, a team led by Enrico Fermi had built the first successful nuclear reactor, which caused a critical chain reaction.

National Archives and Records Administration.

- There are 2 ways to make nuclear fission happen. Rarely, a massive nucleus will simply spontaneously split in half. On the other hand, the process of neutron-induced fission occurs when a fissionable isotope is bombarded with neutrons.

- **Fissionable isotopes** are isotopes that will undergo fission when they're struck with neutrons, sometimes with very high energy, and they include uranium-235, uranium-238, and most other heavy nuclei.

- A **fissile isotope** is one that will undergo fission even if you strike it with a low-energy neutron, including uranium-233, uranium-235, and plutonium-239.

- In neutron-induced fission, a uranium-235 nucleus is bombarded with a neutron, which gets absorbed and delivers some energy to the nucleus. The nucleus becomes uranium-236—because it's still uranium but has picked up an extra neutron—and undergoes an oscillation. It becomes unstable.

- There are many reactions for nuclear fission. Basically, however, a neutron enters the nucleus, the uranium splits into 2 middle-weight nuclei with roughly equal but slightly different masses, and some neutrons leave the nucleus.

- The reaction for neutron-induced fission is $^{1}_{0}n + ^{235}_{92}U \rightarrow X + Y + b\,^{1}_{0}n,$ in which X and Y are the fission products (the middle-weight nuclei) and b is the number, typically between 2 and 3, of additional neutrons.

- In addition to neutrons and middle-weight nuclei, fission energy is released through the process of nuclear fission—in the amount of about 200 million electron volts (eV).

- Nuclear energy is a totally different kind of energy. It is more concentrated than the energy released in a chemical reaction by a factor of about 10 million, which is why nuclear fuels are such concentrated sources of energy.

- It is fundamental to the nature of fission products that they are unstable; they have too many neutrons, so they undergo beta decay. Fission products are intensely radioactive.

- Fission products have half-lives from seconds to typically decades—maybe even centuries in a few cases. Overall, their half-lives are pretty short, so these products must be intensely radioactive to decay that quickly.

334

- We extract energy from nuclear reactions by creating a chain reaction in which many neutrons are released to cause more and more fission until the entire mass of fissile material has undergone fission. This process requires a critical mass of nuclear fuel.

- In fact, if you have a solid mass of plutonium or uranium, it will fission in about a microsecond and release the energy equivalent to thousands and thousands of tons of coal. That's how you create a nuclear fission bomb.

- The likelihood of fission occurring depends on which isotope you're working with. uranium-235 is extremely likely to fission if it's hit by a slow neutron, but it is nearly impossible for uranium-238 to fission if hit by slow neutrons.

- There are many designs for nuclear fission reactors throughout the world, but the most common ones use ordinary water as both the moderator and the coolant that carries the heat away from the fission reaction.

- Fusion is what powers the stars, and it's also responsible for making most of the elements beyond hydrogen and helium. It is used in thermonuclear weapons, but it hasn't been harnessed yet for energy generation on Earth.

- We are still working on harnessing fusion as a practical energy source. If we succeed, each gallon of seawater will become the energy equivalent of about 300 gallons of gasoline, and we would have unlimited energy that would last for 25 billion years.

Important Terms

curve of binding energy: A graph describing the energy release possible in forming atomic nuclei; this graph shows that both fusion of light nuclei and fission of heavy nuclei can release energy.

fissile isotope: An isotope that will undergo fission even if you strike it with a low-energy neutron. Examples include uranium-233, uranium-235, and plutonium-239.

fission: The splitting, spontaneous or induced, of an atomic nucleus into 2 roughly equal pieces. For heavy nuclei (those containing more than 56 protons and neutrons), fission releases large amounts of energy. Nuclear power plants and submarines operate via controlled nuclear fission.

fissionable isotope: An isotope that will undergo fission when it is struck with neutrons, sometimes with very high energy. Examples include uranium-235, uranium-238, and most other heavy nuclei.

fusion: A nuclear reaction in which light nuclei join to produce a heavier nucleus, releasing energy in the process.

Suggested Reading

Rex and Wolfson, *ECP*, chap 25.5–25.6.

Wolfson, *EUP*, chap 38.4–38.5.

Questions to Consider

1. Distinguish nuclear fission and fusion.

2. What is special about fissile isotopes, in contrast to the more general class of fissionable isotopes?

3. What is the role of the moderator in a nuclear reactor?

4. What is a critical mass?

Lecture 57: Energy from the Nucleus

The Particle Zoo
Lecture 58

M odern quantum theory describes the force between 2 particles in terms of their exchanging of a third particle. Through the mid-20th century, physics began to discover more and more seemingly elementary particles. An organizational scheme emerged in which many particles—including the proton and neutron—were seen to be composed of smaller particles called quarks. Today, the standard model of particles and forces describes a universe of fundamental particles. New particle accelerators may yield the answers to unsolved problems in fundamental particle physics.

- In 1932, Sir James Chadwick discovered the neutron. Up until that time, it wasn't absolutely certain what the nucleus was made of— although it was believed that there were 2 kinds of particles in the nucleus, one of them positively charged and one of them neutral.

- In 1932, Carl D. Anderson discovered the positron, which is the antielectron, or the antimatter part of the electron. This is the experimental confirmation of Dirac's 1928 idea of antimatter.

- In 1936, Anderson discovered the particle called the muon, and from 1940 to 1980, there were over 100 "elementary" particles— most of them were not really elementary.

- The final decades of the 20th century consisted of trying to make sense of these many elementary particles and trying to work them into a simpler scheme, which was largely successful, although it is still not known whether the current theory of elementary particles is really solid.

- In 1964 to 1995, that simplification process began with the idea of quarks, particles that made up what were originally thought to be elementary particles. By now, there is solid confirmation that quarks exist, and we understand quite well how they interact.

- There is a standard model of particles and forces that describes the elementary particles as they exist and how they interact. Today, there is a lot of exciting work in high-energy physics that focuses on looking for physics beyond this model.

- There are 2 kinds of elementary particles. There are the particles that make up all ordinary matter, which are called matter particles. They are **fermions**, have 1/2-integer spin, and obey the exclusion principle.

- Ordinary matter comprises less than 5% of the composition of the universe; not much is known about the rest—dark energy and dark matter.

- Then, there are field particles, called **gauge bosons**, that mediate the fundamental forces, such as electromagnetic fields. Bosons have integer spin and don't obey the exclusion principle. Photons are an example.

- The elementary matter particles come in basically 2 kinds: leptons and quarks.

- **Leptons**, or light particles, do not experience the strong force that binds quarks together to make other particles

- Leptons include the electron; the muon, which is similar to the electron but is more massive; and the tau particle, which is even more massive than the muon. They also include 3 kinds of neutrinos: electron neutrinos, muon neutrinos, and tau neutrinos.

- Neutrinos are very light, having almost zero mass. They have very weak interactions with matter and are very difficult to detect.

- Almost every one of these particles has an antiparticle. The electron has the positron. The muon has the antimuon. The tau has the antitau, and there are antineutrinos. There are a few rare exceptions of particles that are their own antiparticles; one example is the photon.

- In addition to leptons are quarks. These are more massive particles, and they come in pairs. There is an up quark, a down quark, a strange quark, a charmed quark, a top quark, and a bottom quark.

- Quarks experience the strong force. They're the building blocks of particles called **hadrons**, which are heavy particles. Protons and neutrons are hadrons, and there are plenty of other hadrons.

- The reason we don't see all these other particles is because almost all of them are unstable; that is, left to their own devices, they quickly decay.

- There are 2 kinds of hadrons: **baryons** and **mesons**.

  ○ Baryons are made up of 3 quarks, and they are fermions with 1/2-integer spin that abide by the exclusion principle. Baryons include the proton and neutron.

  ○ Mesons consist of 2 quarks—a quark and an antiquark. They are bosons because their spin angular momentum cancels so that they have integer spin. They can pile up in the same quantum state, unlike fermions.

- There are 4 particles—8 particles if you count their antiparticles— that constitute a family of elementary particles: the electron, the electron neutrino, the up quark, and the down quark.

- Together, these particles—particularly the electron and the up and down quarks—comprise all ordinary matter. The other particles play a minor role in comprising ordinary matter.

- The second family of these elementary matter particles is composed of the muon, the muon neutrino, the strange quark, and the charmed quark. There's also a third family that contains the tau, the tau neutrino, the bottom quark, and the top quark.

- Physicists aren't going to discover any additional families. There is strong evidence that the number of families is 2.99 with some small range of error. In other words, there are 3 families.

- Composite particles are created out of reactions between these elementary particles. The proton (p) has the charge $+e$, the opposite of the electron's charge. We are pretty sure that the proton is stable; in other words, it lasts forever. The proton's composition is 2 up quarks and 1 down quark.

- The neutron (n) is neutral, and its mass is almost exactly the same as a proton. Its charge is zero. Outside the nucleus, it's unstable with a lifetime of about 20 minutes. The neutron is composed of 1 up quark and 2 down quarks.

- There are some quantities that we know are conserved with elementary particles. Electric charge is conserved, and spin angular momentum is conserved because angular momentum is conserved.

- Not only are there 6 kinds of quarks and their antiquarks, there are 3 kinds of charge called color charge—rather than plus and minus for electric charge—that seem to be conserved.

- In big particle accelerators, 2 protons collide and spew a huge number of particles out of the enormous 7 TeV proton-proton collision energy they have. This collision tells us a lot about protons and the interactions of particles.

- Forces between matter particles, like the electric force or the strong force, result from the exchange of these other particles called gauge bosons. In the electromagnetic case, that's the photon. It has zero mass, and that is what gives the electric force its relatively long range, its $1/r^2$ falloff.

- In the case of the strong force, the particle is called a gluon. There are 8 kinds, and they have zero mass. In the case of the weak force, there are bosons that have mass, and consequently, their force is much weaker.

- In the strong case, it is the fact the photons have charge that makes the math of particle interactions involving the strong force very difficult.

- In addition, the strong force binds quarks together, and we do not think that we could ever break the strong force and pull isolated quarks apart. We don't think we'll ever find an isolated quark.

- As particle accelerators are able to go to higher energies, we're hoping to discover information about the earliest universe because in these massive collisions, we are simulating the very high temperatures and high energies of particles in the earliest instance of the universe.

Important Terms

baryon: Any member of the class of subatomic particles consisting of 3 quarks bound together; protons and neutrons are the most common baryons in the universe today.

fermion: A matter particle, as opposed to a force particle (boson). Fermions take up space and can't be piled on top of each other. Examples include all varieties of quarks and leptons. The spin of a fermion is always a 1/2-integer.

gauge boson: A force-carrying particle, as opposed to a matter particle (fermion). Bosons can be piled on top of each other without limit. Examples include photons, gluons, gravitons, weak bosons, and the Higgs boson. The spin of a boson is always an integer.

hadron: Any particle consisting of quarks bound together by the strong nuclear force. In the universe today, there are 2 families of hadrons.

lepton: Along with quarks, one of the 2 families of particles that represent the current limit on our knowledge of the structure of matter at the smallest scales. The electron, muon, and tau particles, along with their antiparticles and their associated neutrinos, comprise the lepton family; only the electron is stable under the conditions present in the universe today.

meson: A composite bosonic particle consisting of one quark and one antiquark.

Suggested Reading

Rex and Wolfson, *ECP*, chap 26.1–26.4.

Wolfson, *EUP*, chap 39.3–39.4.

Questions to Consider

1. Name some essential differences between leptons and quarks.

2. Describe the difference between baryons and mesons, in terms of quark composition.

3. Why are we unlikely to observe an isolated quark?

Lecture 58: The Particle Zoo

An Evolving Universe
Lecture 59

The belief of modern cosmology is that the universe, at least the part of it we know, began about 14 billion years ago in a colossal explosion called the big bang. The cosmic expansion that started with the big bang continues today. Cosmic expansion was first discovered observationally in the 1920s. The 1965 discovery of the cosmic microwave background—a relic of the time a mere 380,000 years after the big bang—ended any serious doubt about the veracity of the big bang idea.

- Around 1900, there was a widespread view that the universe was a static place—that it didn't evolve overall. Individual stars, planets, and life evolved on Earth, but the universe didn't.

- That belief was so embedded in the scientific culture that Einstein, in 1917, added a so-called cosmological constant to his theory of general relativity.

- One of the first things that came out of general relativity was the idea that the universe must be either expanding or contracting, but by adding this term and adjusting its value, Einstein was able to keep his universe, as predicted by general relativity, static.

- In the 1920s, Edward Hubble discovered cosmic expansion, the first of 3 major discoveries about the universe, by using a new 100-inch telescope on Mount Wilson in California to study distant galaxies.

- In the 1930s and 1940s, as a result of this cosmic expansion idea, there emerged the **big bang theory**, which is the idea that the universe began in some kind of cosmic explosion billions of years ago and has been evolving ever since as the outcome of that explosion.

- In 1948, Sir Hermann Bondi, Thomas Gold, and Fred Hoyle proposed a competing theory called the **steady-state theory**. They managed to preserve a universe that didn't evolve with the theory that nevertheless was consistent with cosmic expansion.

- The result of preserving that consistency was the idea that new matter needed to be created out of nothing in order to populate the universe to the same density as it continued to expand.

- In 1965 came an observation that swept away the possibility of a steady-state universe, and that was the discovery by Arno Penzias and Robert Wilson of the cosmic microwave background radiation.

- Around the 1970s, we began to understand how all the information we were learning about particles fit together and how there existed simpler particles that interacted. Particle physics began to inform cosmology, the study of the universe, at large.

- In 1980, an important idea arose from Alan Guth at MIT. He proposed the idea of cosmic inflation, which solved some problems that arose with the big bang theory of cosmic evolution.

- The big surprise the universe threw at us came in 1998 with the discovery that the universe, which was expected to be slowing down, was actually accelerating.

- The era of precision cosmology enters at around the year 2000 and beyond, in which we can finally measure quantities that refer to the entire universe with the kind of precision we use to measure many other physical properties to several decimal points.

- Hubble began looking at the light from distant galaxies and studying the spectra of that light; he found that the light was redshifted.

- In the Doppler effect, when a source of waves is moving away from you, the waves are stretched out to longer wavelengths. If those waves are waves of light, the light is redshifted.

- The implication of Hubble's finding is that galaxies are receding from us. Furthermore, with his measurements he discovered that the galaxies are receding with speeds that depend on how far away they are: The farther away a galaxy is the more rapidly it's receding from us.

- **Hubble's law** is the simple mathematical relationship that describes that phenomenon: $v = H_0 d$, where H_0 is called the Hubble constant, which is the basic parameter that describes the expansion of the universe.

- The implications of this discovery are that the universe is expanding and that we're not at the center of the universe.

- The universe is expanding, and the only way it could not also be evolving is if somehow new matter were being created to keep its density constant, but that is basically impossible.

- The Hubble constant has the value of about 22 km/s per million light-years. If you were to use Hubble's law to estimate the time since the big bang occurred—in other words, the age of the universe—you would find that number to be 13.6 billion years.

Our own system of stars, the Milky Way Galaxy, emits microwave radiation that can be detected with the aid of radio and infrared telescopes.

- The second of the 3 major discoveries is the 1965 discovery of a bath of microwave radiation that permeates the universe and comes from all directions: the **cosmic microwave background**.

- Arno Penzias and Robert Wilson discovered the remnant of fossil radiation leftover from—not the beginning of the universe, but 380,000 years after the beginning of the universe—the time when atoms were forming.

- Measurements showed that the microwave radiation they discovered, if you plotted the radiance versus wavelength, fit a blackbody curve perfectly—like the curves of hot, glowing objects.

- The entire universe is at a temperature of about 2.7°C above absolute zero, but nevertheless, it's glowing. This glow is the remnant of the radiation that was emitted when atoms were first formed.

- The energy of the radiation was much higher at the time when atoms were formed, but it's been redshifted down to much longer wavelengths by the expansion of the universe. This agrees well with the big bang prediction at the time of the formation of hydrogen atoms.

- The third of the major discoveries is an interaction of cosmology with particle physics: Particle physics is informing cosmology, and cosmology is also informing particle physics.

- Deuterium, helium, lithium, and beryllium were the only elements formed in the big bang. All the rest of the elements were formed in stars by nuclear fusion. The abundances of these elements are very sensitive tests of early universe theories.

- Particle physics helps to confirm our idea that the universe is undergoing expansion following the big bang. What is known about particle physics—and how the particles interact and merge—is completely consistent with what we observe.

- Another idea that arose in 1980 was an idea that was needed to overcome some problems with the big bang theory. This is the idea of the inflationary universe, which was posed by Alan Guth.

- The inflationary universe, a refinement of the big bang theory, blends with our knowledge of particle physics to trace the evolution of the universe from when it was only a fraction of a second old to the present.

- In 1998, cosmic acceleration was discovered. The observation showed that distant supernovas, exploding stars that give out a constant amount of energy, were dimmer than expected. Therefore, the universe is undergoing an accelerated expansion that probably began about 5 billion years ago.

- The big picture of cosmic evolution is that the universe expands, and as it does, it cools. As it cools, more and more particles are able to stick together. As the universe cools, atoms form molecules, molecules form life, and life forms consciousness.

- We now have a classic picture from the WMAP satellite of the cosmic microwave background. Today, our measurements of the cosmic microwave background make cosmology, for the first time, a precision science.

Important Terms

big bang theory: A mathematical solution to the theory of general relativity that implies the universe emerged from an enormously dense and hot state about 13.7 billion years ago.

cosmic microwave background: Electromagnetic radiation in the microwave region of the spectrum, which pervades the universe and represents a "fossil" relic of the time when atoms first formed, about half a million years after the big bang.

Hubble's law: The proportionality between the distance and the apparent recession velocities of galaxies is known as Hubble's law: $v = H_0 d$, where H_0 is called the Hubble constant, the ratio of the speed to the distance. The farther away a galaxy is, the faster it appears to be receding. Hubble's law doesn't apply exactly to nearby galaxies or to galaxies that are very far away.

steady-state theory: The idea, now widely discredited, that the overall structure of the universe never changes.

Suggested Reading

Rex and Wolfson, *ECP*, chap 26.5–26.6.

Wolfson, *EUP*, chap 39.4–39.5.

Questions to Consider

1. What was Hubble's evidence that the universe is expanding?

2. In what sense is the cosmic microwave background a "fossil" from the universe at around age 400,000 years?

3. Name 2 quandaries that are resolved by the inflationary universe theory, and explain how the theory resolves them.

Lecture 59: An Evolving Universe

Humble Physics—What We Don't Know
Lecture 60

M̲ultiple observations lead to the conclusion that the universe contains more matter than we can see and that this dark matter cannot be like ordinary matter. The discovery that the expansion of the universe is actually accelerating shows that an unseen dark energy exists. Observations of distant galaxies, supernovae, and the cosmic microwave background prove that less than 5% of the universe is composed of ordinary matter. Another 23% is dark matter, and most is dark energy. It's humbling that we understand only 5% of our universe!

- We don't understand most of physical reality; we simply don't understand what most of the universe is made of. This is a humbling surprise.

- In particular, we have very little idea what **dark matter** is, but we know that it exists. If that isn't bad enough, even dark matter doesn't account for most of the universe. There is also **dark energy**, about which we know even less.

- The idea of dark matter was first proposed in 1933 in connection with galaxy motions; evidence for dark matter has been building ever since.

- Partly from our understanding of the formation of nuclei in the big bang, we know that dark matter cannot be made of the particles—the neutrons and protons, the quarks—that ordinary matter is made of. It's something new and different.

- Dark matter doesn't interact electromagnetically. Therefore, it doesn't emit electromagnetic waves. That's why it's dark and not luminous; it's not part of the visible universe.

- Furthermore, dark matter is transparent to light and other electromagnetic waves, so we can't detect it by its blocking of electromagnetic waves even if it weren't luminous.

- Dark matter doesn't interact via the strong force either, and that means we can't detect dark matter with interactions involving either electromagnetism or the strong force.

- In addition, dark matter has very limited interactions—if it interacts at all—with atomic nuclei. However, it does interact gravitationally because it is, after all, matter, and matter has mass, and mass is the source of gravity.

- It may be that dark matter also interacts via the weak force, but that depends on the type of particles that comprise dark matter, which we simply don't know. If it did interact by the weak force, we would have indirect means of detection.

- The prime candidates for dark matter are particles called WIMPS, weakly interacting massive particles, and they go beyond predictions of the standard model.

- There have been no confirmed detections of dark matter. However, we do have evidence that dark matter exists.

- From observation, we know that galaxies get together in clusters. When we observe galaxies moving around in clusters, their motions are not consistent gravitationally with the amount of visible matter we see and with the amount of invisible dust that we can infer is in these clusters.

- There are also dwarf galaxies in clusters. We see these very dense clusters of galaxies in which collisions are tearing the larger galaxies apart, and the dwarf galaxies seem to be surviving as though they were protected by a cushion of invisible matter—dark matter.

- Gravitational lensing, the bending of light that Einstein first predicted with general relativity, also gives us evidence of unseen matter that is affecting the trajectories of light from distance objects. Galaxy collisions, which we can look at with gravitational lensing, are an example.

- The cosmic microwave background also provides us with evidence because the cosmic microwave background anisotropies—which vary with position in the sky—are consistent with a flat universe.

- In a flat universe, the density is such that the kinetic energy of the matter in the universe is just barely enough to overcome the gravitational attraction of all the matter, and the universe can just barely have enough energy to expand forever.

- What we find in the universe is that the visible matter, the matter that we can see, is far below what is needed to make up the **critical density**: $p_c = (3H_0^2)/(8\pi G)$.

- This critical density is a really crucial number at understanding how the actual density of the universe compares with the critical density.

- In fact, what these cosmic microwave background observations are telling us is that the density of the universe is essentially the critical density.

- The average density of the entire universe is 10^{-26} kg/m^3, which seems like a very small density, and it is, but it's the average density—including the vast spaces in our Solar System that are basically vacuum.

- What's interesting is that this number, although it's small, is far too big to be explained by the presence of ordinary visible matter.

- If the density is right at the critical density, we have zero total energy. In general relativity, this corresponds to a universe whose overall structure is flat.

- The universe is, in fact, close to the critical density, which is a concept that is summarized by the parameter Ω: the ratio of the actual density of the universe to the critical density of the universe.

- From studies of the cosmic microwave background, it seems that we live in a flat universe where Ω is 1.

- In other words, the value of Ω summarizes the entire fate of the universe. The density is the critical density because Ω is the ratio of the density to the critical density, which is 1.

- The visible matter that we can see has much lower density. In fact, it has a density that's less than about 5% of the critical density. What that means is that visible matter can make up at most about 5% of the substance of the universe.

- Much more substance is needed to make up the other 95%, which leads to the conclusion that there has to be some dark matter that is not visible matter.

- According to general relativity, no matter—whether it's visible or dark—can explain cosmic acceleration, which began 5 billion years ago.

- What cosmic acceleration requires is some kind of repulsive gravity. It requires a new form of invisible energy—not matter—and that's called dark energy.

- There was an earlier epoch in the universe, before about 5 billion years ago, when matter dominated—dark matter, probably. Currently, dark energy dominates. It's the dominant mass-energy in the universe, and we know this from cosmic microwave background observations.

- There's a special kind of energy associated with the vacuum. There must be energy in the vacuum; it has to have a minimum energy that exerts a negative pressure and that gives rise to a repulsion in Einstein's gravitational equations.

- Even though we don't understand dark energy fully, we do know that it has a negative pressure, and therefore, it causes gravity to become repulsive.

- Dark energy acts as a kind of antigravity to augment the expansion of the universe that started with the big bang.

- There's still much more that we don't understand about our universe. We still don't know how to reconcile general relativity—our theory of gravity—with quantum physics. String theory is one such attempt, but the validity and verifiability of string theories remains uncertain.

- In addition, we don't know what, if anything, was before the big bang—or if, perhaps, we occupy one "bubble" in an infinite multiverse of multiple universes, each with its own laws of physics.

- There's a lot to know, and a lot of physics to do, before we reach— if we ever do—a true theory of everything.

Important Terms

critical density: The density that would be required to make the universe spatially flat. Observations of the cosmic microwave background tell us that, in fact, the density of the universe is essentially the critical density.

dark energy: A kind of unseen energy, nature unknown, that drives the accelerating expansion of the universe. One theory is that dark energy is the energy of the quantum vacuum.

dark matter: Matter, not yet seen in laboratory experiments, that is distinct from ordinary matter and would explain the observed lifetimes of galaxies and the rates of rotation of stars in galaxies. Dark matter is thought to comprise most of the universe.

Suggested Reading

Wolfson, *EUP*, chap 39.5.

Questions to Consider

1. Give 2 pieces of evidence for dark matter.

2. Why is dark energy, as distinct from dark matter, needed to explain the acceleration of cosmic expansion?

3. The density of the universe appears to be the critical density. What does this tell us about the overall geometry of the universe?

Glossary

2-slit interference: A process whereby incoming waves of light go through 2 slits in some barrier, and each slit acts as a new source of circular waves. This process results in the same interference pattern as 2-source interference.

2-source interference: A pattern of wave interference in which 2 sources of waves are pulsing at the same frequency, which causes them to send out wave crests that spread out farther and farther.

4-vector: A vector quantity in 4-dimensional space-time that has 4 components: 1 time component and 3 space components.

aberration of starlight: A phenomenon whereby a telescope must be pointed in slightly different directions at different times of year because of Earth's orbital motion. The fact of aberration shows that Earth cannot drag with it the ether in its immediate vicinity and, thus, helps dispel the notion that ether exists.

absolute zero: The absolute limit of cold, at which all heat energy is removed from a system; equal to about −273°C.

acceleration: The rate of change of velocity, measured as distance divided by time.

adiabatic process: A process that takes place without any exchange of heat with its surroundings, during which entropy remains constant. If a gas undergoes adiabatic expansion, its temperature decreases.

alternating current (AC): Electrons in a circuit oscillate back and forth instead of flowing. (Compared with DC, or direct current.)

amplitude: The size of the disturbance that constitutes a wave.

angle of incidence: The perpendicular angle at which a ray enters a system.

angular acceleration: The rate of change of angular velocity.

angular displacement: Rotational analog of change of position.

angular velocity: A measure of the rotation rate of a rotating object.

aperture: Any kind of hole that light can pass through and at which diffraction can occur.

apparent weightlessness: The condition encountered in any freely falling reference frame, such as an orbiting spacecraft, in which all objects have the same acceleration and, thus, seem weightless relative to their local environment.

Archimedes's principle: Discovered in ancient times by the Greek mathematician Archimedes, this principle says that the buoyancy force on a submerged object equals the weight of the displaced fluid.

atomic mass: The sum of the number of nucleons (protons plus neutrons) in the nucleus of an atom (or of all the nucleons in a molecule). The atomic mass number is written to the upper left of the chemical symbol for the element. In all nuclear transformations, the atomic mass is conserved.

atomic number: The total number of protons in an atom's nucleus and, hence, the number of electrons in a neutral atom. Determines what element an atom belongs to.

baryon: Any member of the class of subatomic particles consisting of 3 quarks bound together; protons and neutrons are the most common baryons in the universe today.

Bernoulli's theorem: A statement of energy conservation in a fluid, showing that the pressure is lowest where the flow speed is greatest and vice versa.

big bang theory: A mathematical solution to the theory of general relativity that implies the universe emerged from an enormously dense and hot state about 13.7 billion years ago.

Biot–Savart law: States that a very short length of current produces a magnetic field that falls off as $1/r^2$: $\Delta B = \dfrac{\mu_0}{4\pi}\dfrac{I\Delta Lr}{r^2}$.

Bohr atomic model: The atomic model proposed by Niels Bohr in 1913 in which electrons can only move in discrete orbits around the nucleus. When light is absorbed or emitted, the electron "jumps" from one orbit to another.

Boltzmann's constant: A conversion between temperature and energy. In SI units, it is 1.3×10^{-23} J/K.

buoyancy: The upward force on an object that is less dense than the surrounding fluid, resulting from greater pressure at the bottom of the object.

capacitance: The measure of how much charge a capacitor can hold.

capacitive reactance: When a capacitor acts like a resistor, the resistance is one divided by the quantity frequency times capacitance: $X_C = 1/\omega C$.

capacitor: An energy-storage device that consists of a pair of electrical conductors whose charges are equal but opposite.

Carnot engine: A simple engine that extracts energy from a hot medium and produces useful work. Its efficiency, which is less than 100%, is the highest possible for any heat engine.

Carnot's theorem: A theorem named after French scientist Sadi Carnot that states it is thermodynamically impossible to build an engine whose efficiency is better than a Carnot engine.

center of gravity: For the purposes of the torque that gravity exerts on the object, the point at which an object's mass acts as if all the object's mass were concentrated.

center of mass: An average position of matter in an object; the effective point where gravity (or external force) acts.

centripetal acceleration: The acceleration of an object around any other object or position.

coefficient of thermal expansion: The fractional change that an object undergoes as a result of a temperature change of 1 degree.

collision: An intense interaction between objects that lasts a short time and involves very large forces.

Compton effect: An interaction between a photon and an electron in which the photon scatters off the electron and comes off with less energy. The effect provides a convincing demonstration of the quantization of light energy.

concave mirror: This type of curved mirror is a device that reflects light, forming an inverted real image that is in front of the mirror.

conduction: Heat transfer by physical contact.

conservation of momentum: The situation that exists when the momentum remains unchanged during an interaction.

constant acceleration: Acceleration that increases by the same amount over time.

continuity equation: A statement of mass conservation in a steady flow, stating that the product of density, area, and speed (a quantity expressed in kilograms per second) is constant along the flow tube.

contour line: A line of constant elevation on a map that is perpendicular to the steepest slope.

convection: Heat transfer resulting from fluid motion.

converging lens: A type of lens that sends parallel rays to a focus.

convex lens: A lens that bends outward by taking parallel rays and bending them to a focal point.

Copenhagen interpretation: The standard view of the meaning of quantum physics, which states that it makes no sense to talk about quantities—such as the precise velocity and position of a particle—that cannot, even in principle, be measured simultaneously.

correspondence principle: A principle formulated by Niels Bohr that says quantum mechanics agrees with classical physics, but only in the limit of very large quantum numbers.

cosmic microwave background: Electromagnetic radiation in the microwave region of the spectrum, which pervades the universe and represents a "fossil" relic of the time when atoms first formed, about half a million years after the big bang.

cosmology: The study of the overall structure and evolution of the universe.

Coulomb's law: An equation that predicts the force between any 2 stationary charges at a given distance: $F = \dfrac{kq_1q_2}{r^2}$.

critical density: The density that would be required to make the universe spatially flat. Observations of the cosmic microwave background tell us that, in fact, the density of the universe is essentially the critical density.

curve of binding energy: A graph describing the energy release possible in forming atomic nuclei; this graph shows that both fusion of light nuclei and fission of heavy nuclei can release energy.

damping: The process by which simple harmonic motions tend to lose energy.

dark energy: A kind of unseen energy, nature unknown, that drives the accelerating expansion of the universe. One theory is that dark energy is the energy of the quantum vacuum.

dark matter: Matter, not yet seen in laboratory experiments, that is distinct from ordinary matter and would explain the observed lifetimes of galaxies and the rates of rotation of stars in galaxies. Dark matter is thought to comprise most of the universe.

degenerate matter: Matter that is so tightly crammed together that basically all the particles act as one.

density: The mass per unit volume of a fluid. Its symbol is the Greek letter rho, ρ, and its SI unit is kilograms per cubic meter.

determinism: The belief that future events are completely determined by the present state of the universe—that is, by the exact positions and momenta of all of its particles.

diamagnetism: The opposite of paramagnetism; it's a weak interaction, but it's a repulsive interaction. It occurs when a magnetic field changes near the atomic dipoles, and they respond by developing a magnetic dipole moment that causes them to be repealed from magnets.

diffraction: The phenomenon whereby waves change direction as they go around objects.

diffraction limit: A fundamental limitation posed by the wave nature of light, whereby it is impossible to image an object whose size is smaller than the wavelength of the light being used to observe it.

direct current (DC): Electrons in a circuit flow in only one direction. (Compared with AC, or alternating current.) DC would result from a circuit with a battery; AC would result in household circuits.

displacement: The net change in position of an object from its initial to ending position.

distance: How far apart 2 objects are.

diverging lens: A type of lens that sends parallel rays away from a focus and can only form virtual images.

Doppler effect: Named after a 19th-century Austrian physicist, this is the effect produced when the source of a wave and the observer of the wave are in relative motion. When the 2 are approaching each other, the wavelengths of the wave are compressed, leading to a higher pitch (in sound) or a bluer color (in light). When the 2 are receding, the distance between the wave crests is lengthened, leading to lower pitch or redder light.

eddy current: A current in conductive material caused by changing magnetic fields that dissipates rotational kinetic energy.

elastic collision: A collision in which energy is conserved.

elastic potential energy: The energy that is stored when stretching an object (a spring, for example), which can be measured with the equation $\Delta U_{elastic} = (1/2)kx^2$.

electrical conductor: A material that contains electric charges that are free to move and can, thus, carry electric current. A conductor in which it takes very little energy to promote an electron to a new unoccupied state.

electrical insulator: An insulator that has completely occupied bands—and an energy gap before there are any unoccupied states—so it takes a large amount of energy to promote electrons into the unoccupied states so that they can conduct.

electric charge: The conserved quantity that acts as a source for the electric field.

electric circuit: An electrically conducting path that can carry current in a loop.

electric current: A net flow of electric charge.

electric dipole: A charge distribution that is composed of 2 point charges of equal magnitude but opposite signs.

electric field: The influence that surrounds an electric charge, resulting in forces on other charges.

electric motor: A current loop that is placed in a magnetic field—typically between the poles of a magnet—and rotates on bearings on a shaft.

electric potential difference: The work per unit charge needed to move charge between 2 points, A and B: $\Delta V_{AB} = E\Delta x$.

electric power: The rate of producing or expending energy. In electrical devices, power is the product of voltage and current.

electromagnetic wave: A structure consisting of electric and magnetic fields in which each kind of field generates the other to keep the structure propagating through empty space at the speed of light, c. Electromagnetic waves include radio and TV signals, infrared radiation, visible light, ultraviolet light, X-rays, and gamma rays.

electromagnetism: One of the 4 fundamental forces of nature that involves the interaction of particles having the property of charge; like charges repel, and unlike charges attract. Electromagnetic forces govern the behavior of matter from the scale of the atom to the scale of mountains.

ellipse: A particular kind of stretched circle; the path of planets in orbit.

emergent property: A higher-level property that arises from the micro-level interactions in a complex system.

emf: A term that stands for electromotive force and is a source of electrical energy that maintains a constant potential difference across 2 electrical terminals in a circuit.

energy: The ability to do work.

entropy: A quantitative measure of disorder. The second law of thermodynamics states that the entropy of a closed system can never decrease.

equipotential: Just as contour lines on a map are at right angles to the steepest slope, these lines are at right angles to the electric field.

event: A point in space-time, designated by its location in both space and time.

Faraday's law: A field equation for electromagnetism that describes how electric fields curl around changing magnetic fields (electromagnetic induction).

fermion: A matter particle, as opposed to a force particle (boson). Fermions take up space and can't be piled on top of each other. Examples include all varieties of quarks and leptons. The spin of a fermion is always a 1/2-integer.

ferromagnetism: This is the common, everyday magnetism that is familiar to us. In ferromagnetic materials, there is a very strong interaction among nearby atomic magnetic dipoles that causes them all to align in the same direction.

field line: A visualization tool that is used to picture how electromagnetic fields appear in the presence of sources (charges or currents). A field line shows which way test charges would start to move if released at that point (tangent to, or along, the field lines). Where the field lines bunch together, the forces are strongest.

first law of thermodynamics: The statement that energy is conserved, expanded to include thermal energy.

fissile isotope: An isotope that will undergo fission even if you strike it with a low-energy neutron. Examples include uranium-233, uranium-235, and plutonium-239.

fission: The splitting, spontaneous or induced, of an atomic nucleus into 2 roughly equal pieces. For heavy nuclei (those containing more than 56 protons and neutrons), fission releases large amounts of energy. Nuclear power plants and submarines operate via controlled nuclear fission.

fissionable isotope: An isotope that will undergo fission when it is struck with neutrons, sometimes with very high energy. Examples include uranium-235, uranium-238, and most other heavy nuclei.

fluid dynamics: The study of the behavior of moving fluids.

fluids: Materials that are free to distort and change their shape, such as liquids and gasses.

force: The phenomenon that causes an object to accelerate.

free fall: A state in which gravity is the only force acting on an object.

frequency: The number of wave cycles per second that pass a fixed point in space.

friction: A force that acts between 2 objects in contact, opposing any relative motion between them.

fusion: A nuclear reaction in which light nuclei join to produce a heavier nucleus, releasing energy in the process.

gauge boson: A force-carrying particle, as opposed to a matter particle (fermion). Bosons can be piled on top of each other without limit. Examples include photons, gluons, gravitons, weak bosons, and the Higgs boson. The spin of a boson is always an integer.

Gauss's law: A field equation for electromagnetism that describes how electric fields are produced by electric charges.

generator: A device that uses electromagnetic induction to convert mechanical energy to electrical energy. Typically, a generator involves a coil of wire rotating in a magnetic field.

geocentric: The belief that the Sun and the entire universe rotates around Earth.

geosynchronous orbit: An equatorial orbit at an altitude of about 22,000 miles, where the orbital period is 24 hours. A satellite in such an orbit remains fixed over a point on the equator.

gravitational lensing: An effect caused by the general relativistic bending of light, whereby light from a distant astrophysical object is bent by an intervening massive object to produce multiple and/or distorted images.

gravitational potential energy: The energy content of an object due to its position in relation to other objects, which can be measured with the equation $\Delta U_{\text{gravitational}} = mgh$.

hadron: Any particle consisting of quarks bound together by the strong nuclear force. In the universe today, there are 2 families of hadrons.

heat: The kinetic energy (energy of motion) of the atoms or molecules making up a substance.

heat capacity: The amount of heat energy necessary to increase the temperature of a material by 1°C.

heat of transformation: The energy that it takes to transform from one phase to another; for example, the heat of fusion and the heat of vaporization.

Heisenberg uncertainty principle: The fundamental limit on the precision with which observers can simultaneously measure the position and velocity of a particle. If the position is measured precisely, the velocity will be poorly determined, and vice versa.

heliocentric: The belief that Earth and the rest of the Solar System revolve around the Sun.

Hooke's law: In an ideal spring, the force of the spring is directly proportional to the stretch.

Hubble's law: The proportionality between the distance and the apparent recession velocities of galaxies is known as Hubble's law: $v = H_0 d$, where H_0 is called the Hubble constant, the ratio of the speed to the distance. The farther away a galaxy is, the faster it appears to be receding. Hubble's law doesn't apply exactly to nearby galaxies or to galaxies that are very far away.

Huygens's principle: Named for the Dutch physicist Christiaan Huygens, this principle explains the process whereby each point on a wave crest can be treated as a source of expanding spherical waves, which then interfere to produce propagating waves.

hydrostatic equilibrium: The condition in which there is no net force on a fluid.

ideal gas: A theoretical gas that contains molecules that are far apart and exhibit very few interactions. In the realm where the ideal gas approximation applies, gasses behave basically universally, regardless of the nature of their molecules and regardless of the type of gas.

ideal-gas law: The pressure of a gas times the volume that gas occupies is the product of the number of molecules (the amount of gas), a constant of nature, and the temperature: $pV = NkT$.

incompressible fluids: Fluids for which the density is constant.

inductor: A device that is specifically designed to have a particular value of inductance. Typically, a coil of wire is used as the inductor.

inelastic collision: A collision in which energy is not conserved.

inertial reference frame: A perspective from which a person makes measurements in which Newton's first law holds.

insulating material: A material with no or few free electric charges and, thus, a poor carrier of electric current.

interference: The process whereby 2 waves, occupying the same place at the same time, simply add to produce a composite disturbance. Interference may be constructive, in which the 2 waves reinforce to produce an enhanced composite wave, or destructive, in which case the composite wave is diminished.

interference fringe: Alternating bright and dark bands that are produced by constructive and destructive interference.

invariant: A quantity that has a value that is the same in all frames of reference. The space-time interval is one example of a relativistic invariant.

isothermal process: A process that takes place at a constant temperature. A gas that undergoes isothermal expansion will have to absorb heat from its surroundings.

isotope: Atoms with identical numbers of protons and electrons that differ only in their number of neutrons. Since the number of electrons determines the atom's chemical behavior, all isotopes of an element behave identically in forming molecules with other atoms; the sole difference is their mass.

joule (J): In the International System of Units (SI), the amount of work that can be done by applying a force of 1 newton through a distance of 1 meter. It is named for 19th-century English physicist James Prescott Joule.

kinetic energy: The energy of motion. For a particle of mass m moving with velocity v, the kinetic energy is $K = (1/2)mv^2$.

law of conservation of energy: A fundamental law of physics that states that in a closed system, energy cannot be created or destroyed; it can only change form.

law of inertia: Newton's first law of motion, which states that a body in motion (or at rest) remains in uniform motion (or at rest) unless a force acts on it.

law of reflection: States that a reflected ray goes out at the same angle relative to the perpendicular at which it came into a system.

lens: A piece of transparent material shaped so that refraction brings light rays to a focus.

Lenz's law: States that the direction of any induced voltage or current opposes the change causing it, giving rise to the induced set.

lepton: Along with quarks, one of the 2 families of particles that represent the current limit on our knowledge of the structure of matter at the smallest scales. The electron, muon, and tau particles, along with their antiparticles and their associated neutrinos, comprise the lepton family; only the electron is stable under the conditions present in the universe today.

Lorentz-Fitzgerald contraction: Proposed independently by the Irish physicist Fitzgerald and the Dutch physicist Lorentz, this hypothesis states that when objects move through ether, they are compressed in the direction of motion.

macroscopic properties: A generic term for phenomena and objects at the large scale. Everything that we can directly perceive may be regarded as macroscopic.

macrostate: A state characterized by the number of molecules that are located on each side of a divided area.

magnet: An object that has 2 poles: north and south. North poles repel each other and south poles repel each other, while north attracts south and south attracts north.

magnetic field: The influence surrounding a moving electric charge (and, thus, a magnet) that results in forces on other moving charges (and on magnets or magnetic materials).

magnetic flux: The net amount of a field that flows through a surface. The concept is directly related to flow for the wind field but can be extended by analogy to electric and magnetic fields.

magnetic moment vector: A vector that is perpendicular to the area of a current loop. Its magnitude μ is the current in the loop times the area of the loop.

magnetic monopole: A beginning or end of field lines (e.g., a positive charge is always at the beginning of electric field lines). This concept is important because nobody has ever found a magnetic monopole; therefore, magnetic field lines can never begin or end—they must form loops.

mass: A measure of an object's material content or an object's tendency to resist an acceleration.

meson: A composite bosonic particle consisting of one quark and one antiquark.

metastable equilibrium: An equilibrium that is neither fully stable nor fully unstable.

Michelson-Morley experiment: An experiment performed in the late 19th century by the 2 American physicists from which it takes its name, with the goal of detecting the presence of the ether through which electromagnetic waves moved. Its failure to detect any evidence for the ether led to the development of Einstein's theory of relativity.

microstate: A specific arrangement of individual molecules.

momenergy 4-vector: A vector that combines energy and momentum into one 4-dimensional mathematical vector. Its time component is an object's total energy, heat; its 3 space components are 3 components of momentum.

momentum: The tendency of an object to remain rotating (angular momentum) or to remain in motion in a straight line (linear momentum). Momentum is one of the conserved quantities in nature; in a closed system, it remains unchanged.

net force: The sum of all forces acting on an object.

neutral buoyancy: The state of neither rising nor sinking that occurs for an object of the same density as the surrounding fluid.

neutrino: The lightest of the subatomic particles with masses on the order of 1 millionth of the electron. Neutrinos only interact via the weak nuclear force and, as such, pass through matter with ease. They accompany all beta decays and can be used to peer directly at the nuclear furnace at the core of the Sun.

newton (N): In the International System of Units, the net force required to accelerate a mass of 1 kilogram at a rate of 1 meter per second, squared. It is named for English physicist and mathematician Sir Isaac Newton.

nonconstant acceleration: Acceleration that does not increase by the same amount over time.

nonuniform circular motion: As an object undergoes circular motion, the speed of the object changes.

nucleon: A generic name for neutrons and protons, the constituents of nuclei.

Ohm's law: The statement, valid for some materials, that the electric current is proportional to the applied voltage and inversely proportional to the material's resistance.

optics: The study of light and how light travels through and between materials. (Geometric optics thinks of light as rays; physical optics tends to think of light as waves—both can be important.)

paradigm vector: A vector that describes displacement.

parallel-plate capacitor: A capacitor that contains a pair of parallel conducting plates that are broad in area, as compared to the relatively narrow spacing between them.

paramagnetism: A type of magnetism that is less common than ferromagnetism in which the individual atomic dipoles don't tend to align very strongly.

particle physics: The study of the elementary constituents of nature.

Pauli exclusion principle: The impossibility of putting 2 fermions into the same state. It is this property of fermions that makes them matter particles—they take up space. Bosons, in contrast, do not obey the exclusion principle and can be squeezed together without limit.

period: The time interval between 2 successive wave crests; equivalently, the time for a complete wave cycle.

phase diagram: A diagram showing how the phases of a substance relate to its temperature and pressure.

photoelectric effect: A phenomenon in which light incident on a metal surface causes electrons to be ejected from the surface. The analysis of the photoelectric effect by Albert Einstein was an early success of quantum theory.

photon: The bosonic particle that mediates the electromagnetic force. An electromagnetic wave or field consists of a condensate of a large number of photons. Photons interact directly with any kind of particle that carries electric charge.

Planck's constant: A fundamental constant of nature, designated h (numerically equal to 6.63×10^{-34} J·s), that sets the basic scale of quantization. If h were zero, classical physics would be correct; h being nonzero is what necessitates quantum physics.

position vector: A vector that describes position measured with respect to some origin.

potential energy: The energy content that an object has by virtue of its chemical configuration or its position in space.

power: The rate of producing or expending energy.

precession: The gradual change in direction of a rotating object's rotation axis as a result of an applied torque.

pressure: The force per unit area.

pressure melting: A unique property of water that occurs when the pressure on water in the solid phase is increased, causing it to turn into a liquid as it crosses the solid-liquid boundary.

principle of relativity: A statement that only relative motion is significant. The principle of Galilean relativity is a special case, applicable only to the laws of motion. Einstein's principle of special relativity covers all of physics but is limited to the case of uniform motion.

pV diagram: A diagram in which volume is on the horizontal axis and pressure is on the vertical axis that shows the relationship among pressure, volume, and temperature for an ideal gas.

quantization: This word has a number of meanings in physics. One definition, the one most commonly used, refers to the fact that energy associated with any of the 4 fundamental forces comes in discrete packets—not in a continuum.

quantum tunneling: The surprising phenomenon by which a quantum particle can sometimes pass through a potential energy barrier that would (under classical physics) ordinarily be expected to block it.

quark: One of 6 fundamental particles with fractional charge that combine to make protons and neutrons, among other particles. The types include the up, down, charm, strange, top, and bottom quarks.

radian: The natural measure of angle and, also, the official SI unit of angle; it is the ratio of the arc length to the radius on a circle or circular arc.

radiation: Heat transfer by electromagnetic waves.

radiation pressure: A phenomenon in which the energy from an electromagnetic wave is reflected off the wave when it comes into contact with an object, and the wave obtains twice as much momentum as a result. This phenomenon is expressed as a force per unit area and is the intensity S divided by the speed of light c.

radiocarbon dating: The use of the radioactive isotope of carbon, carbon-14, to determine the age of an object. Carbon-14 is produced in the atmosphere when cosmic rays strike nitrogen atoms in the air.

real image: An image that is small, inverted, and beyond the focal point (where 2 rays meet) in which the eye is seeing light that is actually coming from the image.

reductionism: The philosophical principle that complex systems can be understood once you know what they are made of and how the constituents interact.

reflection: The phenomenon whereby a wave strikes a material and rebounds at the same angle with which it struck the material.

refraction: The phenomenon of waves changing direction of propagation when going from one medium to another.

resistance: The property of a material that describes how it impedes the flow of electric current.

resistor: An element in a circuit formulated to have a specific electrical resistance; it reduces the current that can pass for a given voltage.

resonance: In weakly damped systems, this is the ability to cause large-amplitude oscillations with relatively small force.

rotational inertia: A measure of an object's resistance to change in rotational motion.

rotational motion: Motion about some axis.

scalar: A quantity without direction—just a number.

Schrödinger equation: The equation discovered by Erwin Schrödinger that controls how the quantum wave function behaves over time.

second law of thermodynamics: A general principle stating that systems tend to evolve from more-ordered to less-ordered states.

self-inductance: The property of a circuit that allows the circuit to induce a current in itself.

semiconductor: A material that lies between insulators and conductors in its capacity to carry electric current. The electrical properties of semiconductors are readily manipulated to make the myriad devices at the heart of modern electronics.

shock wave: A very strong, abrupt wave produced when a wave source moves through a medium at a speed faster than the waves in that medium. An example is a sonic boom from a supersonic airplane.

simple harmonic motion (SHM): The motion that occurs when the restoring force or torque is directly proportional to displacement from equilibrium. This motion is characterized by a simple relationship between the position of the object undergoing the motion and the time.

space quantization: A quantization that results from the fact that angular momentum is a vector and that the direction of angular momentum is quantized—as described by the magnetic quantum number m_1: $L = m_1 \hbar$.

space-time: The unification of space and time required by Einstein's theory of relativity as a consequence of the finite and immutable speed of light.

special theory of relativity: Einstein's statement that the laws of physics are the same for all observers in uniform motion.

specific heat: The amount of heat that has to flow into an object for a unit temperature change per unit mass of the object. The SI unit is joules per kilogram per kelvin, or J/kg·K.

speed: The rate of change of position of an object, measured as distance over time.

spherical coordinates: Used to describe the position of a point in terms of its radial distance from some origin and then 2 angles θ, which describe its orientation relative to some axis, and φ, which describes its orientation around that axis.

spin: The intrinsic angular momentum of an elementary particle, which is the property of appearing to have attributes of tiny spinning balls, even though they possess no size at all. The rates of spin of elementary particles are measured in terms of a quantity called h-bar ($\hbar$) and come in any integer or 1/2-integer multiplied by this rate.

stable equilibrium: A system in equilibrium—zero net force and zero net torque—that must also be at a minimum in its potential-energy curve.

standing wave: A wave that "stands" without propagating on a medium of fixed size.

static equilibrium: A state in which an object is subject to zero net force (and zero torque) and does not feel a large or increasing force if it is moved.

steady flow: A special case in which density, pressure, and velocity do not vary with time at a fixed position, although they may vary from position to position.

steady-state theory: The idea, now widely discredited, that the overall structure of the universe never changes.

superconductor: A material that, at sufficiently low temperature, exhibits zero resistance to the flow of electric current.

superfluid: A liquid at extremely low temperatures that has many surprising properties, including zero viscosity.

superposition principle: This principle describes the phenomenon that electric forces add vectorially.

theory: A general principle that is widely accepted and is in accordance with observable facts and experimental data.

theory of everything (TOE): The (as-yet hypothetical) theory that unites all known branches of physics, including classical mechanics, relativity, quantum theory, and so on.

theory of general relativity: Einstein's generalization of special relativity that makes all observers, whatever their states of motion, essentially equivalent. Because of the equivalence principle, general relativity is necessarily a theory about gravity.

thermal expansion: As internal energy increases when heat flows into a material, the mean intermolecular distance increases, resulting in pressure or volume changes.

thought experiment: A highly idealized experiment used to illustrate physical principles.

tidal force: The force that is caused by differences in gravity from place to place; this force is what causes the tides—not gravity.

time dilation: In special relativity, the phenomenon whereby the time measured by a uniformly moving clock present at 2 events is shorter than that measured by separate clocks located at the 2 events. In general relativity, the phenomenon of time running slower in a region of stronger gravity (greater space-time curvature).

torque: The rotational analog of force; torque depends on force and where that force is applied.

total internal reflection: Complete reflection that occurs as light attempts to go from a more dense to a less dense medium, as from water to air.

transformer: A device that uses electromagnetic induction to transform high-voltage/low-current electricity to low-voltage/high-current, and vice versa.

translational motion: Moving from place to place.

transverse wave: A wave that results from a disturbance at right angles, or perpendicular, to the medium in which the wave is propagating.

trigonometry: The branch of mathematics that studies the relationships among the parts of a triangle.

triple point: The point that defines a unique temperature and pressure for a substance at which its phases can coexist. For water, it is the point at which liquid water and solid water, or ice and water vapor, can coexist.

ultraviolet catastrophe: The absurd prediction of classical physics that a hot, glowing object should emit an infinite amount of energy in the short-wavelength region of the electromagnetic spectrum.

uniform circular motion: Circular motion that occurs at a constant speed.

universal gravitation: The concept, originated by Newton, that every piece of matter in the universe attracts every other piece.

vector: A quantity that has both magnitude and direction.

velocity: Average velocity is total distance divided by the time it took to traverse that distance; units are length per time (for example, miles per hour).

velocity vector: A vector that determines the rate of change of position.

virtual image: An image that is not, in some sense, really there because light is not actually coming from the place that the image is.

watt (W): In the International System of Units, the rate of energy conversion equivalent to 1 joule per second. It is named for Scottish engineer James Watt.

wave: A traveling disturbance that carries energy but not matter. Waves may either be traveling (like a moving sound wave) or standing (like the vibrations of a wire with fixed ends).

wavelength: The distance between successive crests of a periodic wave.

weight: The force of the gravitational pull on a mass.

work: The exertion of a force over a distance.

work-energy theorem: The change in kinetic energy is equal to the net work done on an object ($\Delta K = W_{net}$).

Bibliography

Atkins, Peter. *Four Laws That Drive the Universe*. Oxford: Oxford University Press, 2007. A concise exposition of the laws of thermodynamics using minimal mathematics.

Baigrie, Brian. *Electricity and Magnetism: A Historical Perspective*. Westport, CT: Greenwood, 2006. One of the newer of many available histories of electromagnetism.

Davies, Paul, and Niels Gregersen, eds. *Information and the Nature of Reality: From Physics to Metaphysics*. Cambridge: Cambridge University Press, 2010. A provocative compilation of writings by scientists, philosophers, and theologians about the nature of physical reality. Includes discussions of emergent properties. A good ancillary reading both at the beginning and end of *Physics and Our Universe: How It All Works*.

Dolnick, Edward. *The Clockwork Universe*. New York: Harper, 2011. A general history of early Western science and its relation to religion.

Elmore, William, and Mark Heald. *Physics of Waves*. New York: Dover, 1985. A classic undergraduate text that goes deeper into wave motion than this course.

Ford, Kenneth, and Diane Goldstein. *The Quantum World: Quantum Physics for Everyone*. Cambridge, MA: Harvard University Press, 2005. A survey of quantum physics by a former president of the American Physical Society. Goes far beyond the quantum physics covered in this course.

Gleick, James. *Isaac Newton*. New York: Vintage Press, 2004. A new biography by a highly regarded popularizer of science and biographer of scientists.

Greene, Brian. *The Elegant Universe: Superstrings, Hidden Dimensions, and the Quest for the Ultimate Theory*. New York: Vintage Books, 2000. A lively exposition of contemporary fundamental physics, especially string theory, by one of that theory's most eloquent proponents.

Greenier, Robert. *Rainbows, Halos, and Glories*. Milwaukee, WI: Peanut Butter Publishing, 1999. Explanations and beautiful photos of common and rare optical phenomena.

Harman, Peter M. *The Natural Philosophy of James Clerk Maxwell*. Cambridge: Cambridge University Press, 2001. A scholarly look at the evolution of Maxwell's ideas on physics. Not only will you learn about James Clerk Maxwell's culminating theory of electromagnetism, but you will also learn about his substantial contributions to thermodynamics and statistical mechanics.

James, Frank A. J. L. *Michael Faraday: A Very Short Introduction*. Oxford: Oxford University Press, 2010. Electromagnetic induction is one of the most fascinating and useful phenomena in physics. This brief biography covers the life and science of induction's discoverer, Michael Faraday.

Kirshner, Robert. *The Extravagant Universe: Exploding Stars, Dark Energy, and the Accelerating Cosmos*. Princeton, NJ: Princeton University Press, 2004. An account of modern cosmology by an astrophysicist who was involved in the discovery of cosmic acceleration.

Lindley, David. *Uncertainty: Einstein, Bohr, and the Struggle for the Soul of Science*. New York: Anchor, 2008. Lindley—a former editor at *Nature*, *Science*, and *Science News*—chronicles the debates between Einstein and Bohr over the meaning of quantum physics.

———. *Where Does the Weirdness Go? Why Quantum Mechanics Is Strange, but Not as Strange as You Think*. New York: Basic Books, 1996. One of the best explications of some of the weirder phenomena in quantum physics, including entanglement.

Livingston, Dorothy Michelson. *The Master of Light: A Biography of Albert A. Michelson*. Chicago: University of Chicago Press, 1979. Michelson's daughter writes about her father's life and science, his famous experiment, and his skill as an experimental scientist.

Mather, John, and John Boslough. *The Very First Light: The True Inside Story of the Scientific Journey Back to the Dawn of the Universe*. New York: Basic Books, 2008. The history and science of the cosmic microwave background as told by John Mather, Nobel laureate and chief project scientist for the Cosmic Background Explorer satellite.

Newton, Isaac. *Opticks*. Amherst, NY: Prometheus Books, 2003. For a real historical context, try this original book, written by Sir Isaac Newton. Many versions of the English translation are available.

———. *The Principia: Mathematical Principles of Natural Philosophy*. Berkeley: University of California Press, 1999. This original book was written by Sir Isaac Newton and will provide historical context. This is the highly regarded Berkeley edition, possibly the first new translation from the Latin in several hundred years, and it includes a guide to the *Principia*. Earlier translations are also widely available.

Panek, Richard. *The 4% Universe: Dark Matter, Dark Energy, and the Race to Discover the Rest of Reality*. Boston: Houghton Mifflin Harcourt, 2011. A science writer chronicles the story of dark matter and dark energy, which leave us a universe of which only 4% of the contents is the ordinary matter that we understand.

Potter, Merle C. *Fluid Mechanics Demystified*. New York: McGraw-Hill, 2009. Fluid mechanics is a difficult subject, and this book—part of the popular *Demystified* series—helps to demystify it at a higher mathematical level than this course.

Rex, Andrew, and Richard Wolfson. *Essential College Physics*. Boston: Pearson Addison-Wesley, 2010. A general introduction to measurement and quantitative methods in physics.

Taylor, Edwin, and John Archibald Wheeler. *Spacetime Physics*. 2nd ed. San Francisco: W. H. Freeman, 1992. My favorite introduction to special relativity, at a slightly higher mathematical level than this course and with a strong emphasis on the fundamental principles underlying relativity. Supplies great analogies.

Wolfson, Richard. *Essential University Physics*. 2nd ed. Boston: Pearson Addison-Wesley, forthcoming. This edition is scheduled to be released in 2012; the first edition was published in 2007. Covers realms of physics as well as measurement and quantitative methods.

———. *Simply Einstein: Relativity Demystified*. New York: W. W. Norton, 2003. My own presentation of special and general relativity that is more thorough than this course but less mathematical.

Wood, Elizabeth. *Science for the Airline Passenger*. New York: Ballantine, 1969. Take this little book on your next airplane trip; it describes many of the optical phenomena you'll see out the window as well as covering other scientific aspects of your flight. It is out of print, so look for a used copy.